(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 January 2001 (18.01.2001)

PCT

(10) International Publication Number
WO 01/04137 A1

(51) International Patent Classification[7]: C07H 21/02, 21/04, A01K 67/00

(21) International Application Number: PCT/US00/18803

(22) International Filing Date: 10 July 2000 (10.07.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/143,020 9 July 1999 (09.07.1999) US

(71) Applicant *(for all designated States except US)*: GENAISSANCE PHARMACEUTICALS, INC. [US/US]; Five Science Park, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: CHEW, Anne [US/US]; 837 Orange Street, Apartment 2, New Haven, CT 06511 (US). DENTON, R., Rex [US/US]; 129 Hunters Trail, Madison, CT 06405 (US). DUDA, Amy [US/US]; 93 Marvel Road, New Haven, CT 06515 (US). NANDABALAN, Krishnan [IN/US]; 228 Village Pond Road, Guilford, CT 06437 (US). STEPHENS, Joel, Claiborne [US/US]; 46 Crabapple Lane, Guilford, CT 06437 (US).

(74) Agent: HENDERSON, Melodie, W.; Genaissance Pharmaceuticals, Inc., Five Science Park, New Haven, CT 06511 (US).

(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*
— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/04137 A1

(54) Title: DRUG TARGET ISOGENES: POLYMORPHISMS IN THE OSTEOCLASTOGENESIS INHIBITORY FACTOR GENE

(57) Abstract: Polynucleotides comprising one or more of 24 novel single nucleotide polymorphisms in the human Osteoclastogenesis Inhibitory Factor (TNFRSF11B) gene are described. Compositions and methods for detecting one or more of these polymorphisms are also disclosed.In addition, various genotypes and haplotypes for TNFRSF11B gene that exist in the population are described.

DRUG TARGET ISOGENES: POLYMORPHISMS IN THE OSTEOCLASTOGENESIS INHIBITORY FACTOR GENE

RELATED APPLICATIONS

This application is a continuation-in-part of, and claims priority to, U.S. Provisional Application Serial No. 60/143,020 filed July 9, 1999.

FIELD OF THE INVENTION

This invention relates to variation in genes that encode pharmaceutically important proteins. In particular, this invention provides genetic variants of the human Osteoclastogenesis Inhibitory Factor (TNFRSF11B) gene and methods for identifying which variant(s) of this gene is/are possessed by an individual.

BACKGROUND OF THE INVENTION

Current methods for identifying pharmaceuticals to treat disease often start by identifying, cloning, and expressing an important target protein related to the disease. A determination of whether an agonist or antagonist is needed to produce an effect that may benefit a patient with the disease is then made. Then, vast numbers of compounds are screened against the target protein to find new potential drugs. The desired outcome of this process is a drug that is specific for the target, thereby reducing the incidence of the undesired side effects usually caused by a compound's activity at non-intended targets.

What this approach fails to consider, however, is that natural variability exists in any and every population with respect to a particular protein. A target protein currently used to screen drugs typically is expressed by a gene cloned from an individual who was arbitrarily selected. However, the nucleotide sequence of a particular gene may vary tremendously among individuals. Subtle alteration(s) in the primary nucleotide sequence of a gene encoding a target protein may be manifested as significant variation in expression of or in the structure and/or function of the protein. Such alterations may explain the relatively high degree of uncertainty inherent in treatment of individuals with drugs whose design is based upon a single representative example of the target. For example, it is well-established that some classes of drugs frequently have lower efficacy in some individuals than others, which means such individuals and their physicians must weigh the possible benefit of a larger dosage against a greater risk of side effects. In addition, variable information on the biological function or effects of a particular protein may be due to different scientists unknowingly studying different isoforms of the gene encoding the protein. Thus, information on the type and frequency of genomic variation that exists for pharmaceutically important proteins would be useful.

The organization of single nucleotide variations (polymorphisms) in the primary sequence of a gene into one of the limited number of combinations that exist as units of inheritance is termed a haplotype. Each haplotype therefore contains significantly more information than individual unorganized

polymorphisms. Haplotypes provide an accurate measurement of the genomic variation in the two chromosomes of an individual.

It is well-established that many diseases are associated with specific variations in gene sequences. However while there are examples in which individual polymorphisms act as genetic markers for a particular phenotype, in other cases an individual polymorphism may be found in a variety of genomic backgrounds and therefore shows no definitive coupling between the polymorphism and the causative site for the phenotype (Clark AG et al. 1998 *Am J Hum Genet* 63:595-612; Ulbrecht M et al. 2000 *Am J Respir Crit Care Med* 161: 469-74). In addition, the marker may be predictive in some populations, but not in other populations (Clark AG et al. 1998 *supra*). In these instances, a haplotype will provide a superior genetic marker for the phenotype (Clark AG et al. 1998 *supra*; Ulbrecht M et al. 2000, *supra*; Ruaño G & Stephens JC *Gen Eng News* 19 (21), December 1999).

Analysis of the association between each observed haplotype and a particular phenotype permits ranking of each haplotype by its statistical power of prediction for the phenotype. Haplotypes found to be strongly associated with the phenotype can then have that positive association confirmed by alternative methods to minimize false positives. For a gene suspected to be associated with a particular phenotype, if no observed haplotypes for that gene show association with the phenotype of interest, then it may be inferred that variation in the gene has little, if any, involvement with that phenotype (Ruaño & Stephens 1999, *supra*). Thus, information on the observed haplotypes and their frequency of occurrence in various population groups will be useful in a variety of research and clinical applications.

One possible drug target for the treatment of osteoporosis and other disorders caused by abnormal osteoclast recruitment and function is the Osteoclastogenesis Inhibitory Factor (TNFRSF11B) gene or its encoded product. TNFRSF11B is also known as Osteoprotegerin, (OPG) (Simonet et al., *90 Cell* 89(2):309-319, 1997) and TNF receptor-like molecule 1 (TR1) (Takahashi et al., *Endocr. Rev.* 20:345-357, 1999), and is also sometimes referred to by the symbols OCIF or OIF. TNFRSF11B is a soluble member of the tumor necrosis factor receptor (TNFR) superfamily and binds to at least one TNF-related cytokine, RANKL (also known as TRANCE, OPGL, ODF), which stimulates differentiation of osteoclasts, which are the bone resorbing cells in the body, from osteoclast precursors by binding to RANK, a TNF receptor family member (Tsuda E. & Higashio K 1998 *Nippon Rinsho* 56: 1435-9). *In vitro* studies have shown that TNFRSF11B neutralizes RANKL-induced osteoclastogenesis by binding to RANKL, suggesting that TNFRSF11B is actually a secreted "decoy" receptor for RANKL that blocks initiation of a critical RANK-RANKL signal transduction pathway within osteoclast precursor cells (Takahshi et al., *Biochem. Biophys. Res. Commun.* 256:449-455, 1999). As a result of this blocking action, the number of mature osteoclasts is decreased. *In vivo*, TNFRSF11B increases bone mineral density and bone volume in normal rats (*Nippon Rinsho* 56:1435-1439, 1998), and also exhibits hypocalcemic effects in normal mice and in hypercalcemic nude mice carrying tumors associated with humoral hypercalcemia of malignancy (*Bone* 23:495-498, 1998). Also, it was reported that TNFRSF11B knock-out mice develop severe osteoporosis due to enhanced osteoclastogenesis when they grew to be

adults (Mizuno et al., *Biochem. Biophys. Res. Commun.* 247:610-615, 1998). Thus, along with RANKL and RANK, TNFRSF11B is one of the key molecules that regulate osteoclast recruitment and function, and as such, an understanding of variation in the TNFRSF11B gene should be useful in developing new therapies for metabolic diseases caused by abnormal osteoclast recruitment and function such as osteopetrosis, osteoporosis, metastatic bone disease, Paget's disease, rheumatoid arthritis, and periodontal bone disease.

The human, mouse and rat TNFRSF11B proteins are all 401 amino acids in length, with human and rat TNFRSF11B having 94% amino acid (a.a.) sequence identity (Akatsu et al. 1998 *Bone* 23: 495-8). Within the TNFR superfamily, TNFRSF11B is most similar to TNFRII and CD40, in that this secreted protein has no transmembrane segment, and circulates as a disulfide-linked homodimer. TNFRSF11B has four cysteine-rich domains and two death domain homologous regions present in tandem at the C-terminal portion of the protein (Morinaga et al., *Eur. J. Biochem.* 254:6850691, 1998; Mizuno et al., *Gene* 214:339-343, 1998).

Human TNFRSF11B is encoded by a single-copy gene having five exons and four introns which span 29 kilobases (kb) on chromosome 8q24 of the human genome (Morinaga et al., *Eur. J. Biochem.* 254:685-691, 1998). Although the full genomic sequence has not been published, a reference sequence for this gene comprises the partial sequences shown in Fig. 1 (GenBank Accession No. AB008821.1; SEQ ID NO:1), which includes the 5′ untranslated region and the coding sequence for exon 1, and Fig. 2 (GenBank Accession No. AB008822.1; SEQ ID NO:2), which includes the coding sequences for exons 2-5 as well as the 3′ untranslated region. Reference sequences for an TNFRSF11B cDNA and protein are shown in Figures 3 (SEQ ID NO:3) and 4 (SEQ ID NO:4), respectively.

Expression regulatory elements identified in this nearly complete genomic sequence include (1) a major transcription initiation site located 67 nucleotides (nt) upstream of the initiation ATG codon, (2) a translation-termination codon in exon 5, and (3) a typical poly(A)-addition signal located 173 nt downstream of the translation-termination codon (Morinaga et al., *supra*). Two TNFRSF11B transcripts of 4.2 kb and 6.5 kb have been detected in IMR-90 cells, with the shorter transcript containing the 3′ half of intron 2 and the longer transcript containing all of intron 2 (Morinaga et al., *supra*).

One group has reported finding single nucleotide polymorphisms (SNPs) at positions 9 and 22 of exon 1, which correspond to positions 1181 and 1194 of GENBANK Accession #AB008821.1 (Yasuda et al., *Endocrinology* 139:1329-1337, 1997; GENBANK Acc. No. E15271.1). The SNPs at these sites would result in variation in the encoded amino acid sequence at position 3 (lysine or asparagine) and/or position 8 (alanine or serine), respectively, depending on the particular combination of nucleotides found at these polymorphic sites in one of the two copies of the TNFRSF11B gene from an individual.

Because of the potential for polymorphisms in the TNFRSF11B gene to affect the expression and function of the encoded protein, it would be useful to determine whether additional polymorphisms exist in the TNFRSF11B gene, as well as how such polymorphisms are combined in different copies of the gene. Such information would be useful for studying the biological function of TNFRSF11B as well as in

identifying drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

SUMMARY OF THE INVENTION

Accordingly, the inventors herein have discovered 24 novel polymorphic sites in the TNFRSF11B gene. These polymorphic sites (PS) correspond to the following nucleotide positions in the indicated GenBank Accession Number: 491 (PS1), 676 (PS2), 889 (PS3), 916 (PS4), 950 (PS5), 1217 (PS7), 1294 (PS8) and 1390 (PS9) in AB008821; 505 (PS10), 668 (PS11), 4397 (PS12), 4501 (PS13), 6601 (PS14), 6824 (PS15), 6839 (PS16), 6845 (PS17), 6893 (PS18), 6950 (PS19), 8258 (PS20), 8391 (PS21), 8622 (PS22), 8694 (PS23), 8955 (PS24) and 9003 (PS25) in AB008822. The polymorphisms at these sites are guanine or adenine at PS1, guanine or thymine at PS2, cytosine or thymine at PS3, guanine or thymine at PS4, thymine or cytosine at PS5, cytosine or thymine at PS7, guanine or adenine at PS8, cytosine or adenine at PS9, cytosine or thymine at PS10, thymine or cytosine at PS11, thymine or cytosine at PS12, cytosine or thymine at PS13, guanine or adenine at PS14, cytosine or thymine at PS15, guanine or adenine in PS16, adenine or guanine at PS17, adenine or guanine at PS18, adenine or cytosine at PS19, guanine or adenine at PS20, thymine or cytosine at PS21, guanine or thymine at PS22, guanine or adenine at PS23, adenine or thymine at PS24, and thymine or cytosine at PS25. In addition, the inventors confirmed the presence of the previously reported polymorphic site at nucleotide 1181 (PS6) in Figure 1. The polymorphic site at 1194 was not detected in the experiments described herein. It is believed that TNFRSF11B-encoding polynucleotides containing one or more of the novel polymorphic sites reported herein will be useful in studying the expression and biological function of TNFRSF11B, as well as in developing drugs targeting this protein. In addition, information on the combinations of polymorphisms in the TNFRSF11B gene may have diagnostic and forensic applications.

Thus, in one embodiment, the invention provides an isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the TNFRSF11B gene or a fragment thereof. The reference sequence comprises SEQ ID NOS:1-2 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of adenine at PS1, thymine at PS2, thymine at PS3, thymine at PS4, cytosine at PS5, thymine at PS7, adenine at PS8, adenine at PS9, thymine at PS10, cytosine at PS11, cytosine at PS12, thymine at PS13, adenine at PS14, thymine at PS15, adenine at PS16, guanine at PS17, guanine at PS18, cytosine at PS19, adenine at PS20, cytosine at PS21, thymine at PS22, adenine at PS23, thymine at PS24 and cytosine at PS25. In a preferred embodiment, the polymorphic variant comprises an additional polymorphism of cytosine at PS6. A particularly preferred polymorphic variant is a naturally-occurring isoform (also referred to herein as an "isogene") of the TNFRSF11B gene. A TNFRSF11B isogene of the invention comprises guanine or adenine at PS1, guanine or thymine at PS2, cytosine or thymine at PS3, guanine or thymine at PS4, thymine or cytosine at PS5, cytosine or thymine at PS7, guanine or adenine at PS8, cytosine or adenine at PS9, cytosine or thymine at PS10, thymine or cytosine at PS11, thymine or cytosine at PS12, cytosine or thymine at PS13,

guanine or adenine at PS14, cytosine or thymine at PS15, guanine or adenine in PS16, adenine or guanine at PS17, adenine or guanine at PS18, adenine or cytosine at PS19, guanine or adenine at PS20, thymine or cytosine at PS21, guanine or thymine at PS22, guanine or adenine at PS23, adenine or thymine at PS24, and thymine or cytosine at PS25. The invention also provides a collection of TNFRSF11B isogenes, referred to herein as a TNFRSF11B genome anthology.

A TNFRSF11B isogene may be defined by the combination and order of these polymorphisms in the isogene, which is referred to herein as a TNFRSF11B haplotype. Thus, the invention also provides data on the number of different TNFRSF11B haplotypes found in the reference populations used in the experiments described herein. This haplotype data is useful in methods for deriving a TNFRSF11B haplotype from an individual's genotype for the TNFRSF11B gene and for determining an association between a TNFRSF11B haplotype and a particular trait.

In another embodiment, the invention provides a polynucleotide comprising a polymorphic variant of a reference sequence for a TNFRSF11B cDNA or a fragment thereof. The reference sequence comprises SEQ ID NO:3 (Fig. 3) and the polymorphic cDNA comprises at least one polymorphism selected from the group consisting of thymine at a position corresponding to nucleotide 699, adenine at a position corresponding to nucleotide 714, guanine at a position corresponding to nucleotide 720, guanine at a position corresponding to nucleotide 768, adenine at a position corresponding to nucleotide 841, thymine at a position corresponding to nucleotide 1102 and cytosine at a position corresponding to nucleotide 1150. In a preferred embodiment, the polymorphic variant comprises an additional polymorphism of cytosine at a position corresponding to nucleotide 9 in Figure 3.

Polynucleotides complementary to these TNFRSF11B genomic and cDNA variants are also provided by the invention.

In other embodiments, the invention provides a recombinant expression vector comprising one of the polymorphic genomic variants operably linked to expression regulatory elements as well as a recombinant host cell transformed or transfected with the expression vector. The recombinant vector and host cell may be used to express TNFRSF11B for protein structure analysis and drug binding studies.

In yet another embodiment, the invention provides a polypeptide comprising a polymorphic variant of a reference amino acid sequence for the TNFRSF11B protein. The reference amino acid sequence comprises SEQ ID NO:4 (Fig. 4) and the polymorphic variant comprises at least one variant amino acid selected from the group consisting of methionine at a position corresponding to amino acid position 240, methionine at a position corresponding to amino acid position 281, and serine at a position corresponding to amino acid 368. In some embodiments, the polymorphic variant also comprises asparagine at a position corresponding to amino acid 3 in Figure 4. A polymorphic variant of TNFRSF11B is useful in studying the effect of the variation on the biological activity of TNFRSF11B as well as studying the binding affinity of candidate drugs targeting TNFRSF11B for the treatment of osteoporosis and other disorders caused by abnormal osteoclast recruitment and function.

The present invention also provides antibodies that recognize and bind to the above polymorphic

TNFRSF11B protein variant. Such antibodies can be utilized in a variety of diagnostic and prognostic formats and therapeutic methods.

In other embodiments, the invention provides methods, compositions, and kits for haplotyping and/or genotyping the TNFRSF11B gene in an individual. The methods involve identifying the nucleotide or nucleotide pair present at one or more polymorphic sites selected from PS1-5,PS7-25 in one or both copies of the TNFRSF11B gene from the individual. The compositions contain oligonucleotide probes and primers designed to specifically hybridize to one or more target regions containing, or that are adjacent to, a polymorphic site. The methods and compositions for establishing the genotype or haplotype of an individual at the novel polymorphic sites described herein are useful for studying the effect of the polymorphisms in the etiology of diseases affected by the expression and function of the TNFRSF11B protein, studying the efficacy of drugs targeting TNFRSF11B, predicting individual susceptibility to diseases affected by the expression and function of the TNFRSF11B protein and predicting individual responsiveness to drugs targeting TNFRSF11B.

In yet another embodiment, the invention provides a method for identifying an association between a genotype or haplotype and a trait. In preferred embodiments, the trait is susceptibility to a disease, severity of a disease, the staging of a disease or response to a drug. Such methods have applicability in developing diagnostic tests and therapeutic treatments for osteoporosis and other disorders caused by abnormal osteoclast recruitment and function.

The present invention also provides transgenic animals comprising one of the TNFRSF11B genomic polymorphic variants described herein and methods for producing such animals. The transgenic animals are useful for studying expression of the TNFRSF11B isogenes *in vivo*, for *in vivo* screening and testing of drugs targeted against TNFRSF11B protein, and for testing the efficacy of therapeutic agents and compounds for osteoporosis and other disorders caused by abnormal osteoclast recruitment and function in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the TNFRSF11B gene. The computer system comprises a computer processing unit; a display; and a database containing the polymorphism data. The polymorphism data includes the polymorphisms, the genotypes and the haplotypes identified for the TNFRSF11B gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing TNFRSF11B haplotypes organized according to their evolutionary relationships.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a partial reference sequence for the TNFRSF11B gene (Genbank Version Number AB008821.1; contiguous lines; SEQ ID NO:1), with the underline indicating the start codon, shading indicating the reference coding sequence, and bold nucleotides indicating the polymorphic sites and polymorphisms identified by Applicants in a reference population.

Figure 2 illustrates a partial reference sequence for the TNFRSF11B gene (Genbank Version

Number AB008822.1; contiguous lines; SEQ ID NO:2), with the underline indicating the stop codon, shading indicating the reference coding sequence, and bold nucleotides indicating the polymorphic sites and polymorphisms identified by Applicants in a reference population.

Figure 3 illustrates a reference sequence for the TNFRSF11B coding sequence (GENBANK ACC# AB002146; contiguous lines; SEQ ID NO:3) with underlines indicating the start and stop codons, and bold nucleotides indicating the polymorphic sites and polymorphisms identified by Applicants in a reference population.

Figure 4 illustrates a reference sequence for the TNFRSF11B protein (GENBANK ACC # BAA25910; contiguous lines; SEQ ID NO:4) with the bold amino acids indicating the amino acid variations caused by the polymorphisms of Figure 3.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the TNFRSF11B gene. As described in more detail below, the inventors herein discovered 24 novel polymorphic sites by characterizing the TNFRSF11B gene found in genomic DNAs isolated from Index Repository IA that contains immortalized cell lines from one chimpanzee and 93 human individuals and Index Repository IB that contains 70 human individuals. These two repositories contain 51 individuals in common.

The human individuals in Index Repository IA included a reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: Caucasian (22 individuals), African descent (20 individuals) Asian (20 individuals) Hispanic/Latino (17 individuals). To the extent possible, the members of this reference population were organized into population subgroups by the self-identified ethnogeographic origin of their four grandparents as shown in Table 1 below. In addition, Index Repository IA contains three unrelated indigenous American Indians (one from each of North, Central and South America), one three-generation Caucasian family (from the CEPH Utah cohort) and one two-generation African-American family.

Table 1. Population Groups in Index Repository IA

Population Group	Population Subgroup	No. of Individuals
African descent		20
	Sierra Leone	1
Asian		20
	Burma	1
	China	3
	Japan	6
	Korea	1
	Philippines	5
	Vietnam	4
Caucasian		22
	British Isles	3
	British Isles/Central	4
	British Isles/Eastern	1
	Central/Eastern	1
	Eastern	3
	Central/Mediterranean	1
	Mediterranean	2
	Scandinavian	2
Hispanic/Latino		17
	Caribbean	7
	Caribbean (Spanish Descent)	2
	Central American (Spanish Descent)	1
	Mexican American	4
	South American (Spanish Descent)	3

Index Repository IB contains a reference population of 70 human individuals comprised of 4 three-generation families (from the CEPH Utah cohort) as well as unrelated African-American, Asian and Caucasian individuals. A total of 38 individuals in this reference population are unrelated.

Using the TNFRSF11B genotypes identified in the Index Repositories and the methodology described in the Examples below, the inventors herein also determined the haplotypes found on each chromosome for most human members of these repositories. The TNFRSF11B genotypes and haplotypes found in the Index Repositories include those shown in Tables 4 and 5, respectively. The polymorphism and haplotype data disclosed herein are useful for studying population diversity, anthropological lineage, the significance of diversity and lineage at the phenotypic level, paternity testing, forensic applications, and for identifying associations between the TNFRSF11B genetic variation and a trait such as level of drug response or susceptibility to disease.

In the context of this disclosure, the following terms shall be defined as follows unless otherwise indicated:

Allele - A particular form of a genetic locus, distinguished from other forms by its particular nucleotide sequence.

Candidate Gene – A gene which is hypothesized to be responsible for a disease, condition, or the response to a treatment, or to be correlated with one of these.

Gene - A segment of DNA that contains all the information for the regulated biosynthesis of an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

Genotype – An unphased 5′ to 3′ sequence of nucleotide pair(s) found at one or more polymorphic sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype as described below.

Full-genotype – The unphased 5′ to 3′ sequence of nucleotide pairs found at all known polymorphic sites in a locus on a pair of homologous chromosomes in a single individual.

Sub-genotype – The unphased 5′ to 3′ sequence of nucleotides seen at a subset of the known polymorphic sites in a locus on a pair of homologous chromosomes in a single individual.

Genotyping – A process for determining a genotype of an individual.

Haplotype – A 5′ to 3′ sequence of nucleotides found at one or more polymorphic sites in a locus on a single chromosome from a single individual. As used herein, haplotype includes a full-haplotype and/or a sub-haplotype as described below.

Full-haplotype – The 5′ to 3′ sequence of nucleotides found at all known polymorphic sites in a locus on a single chromosome from a single individual.

Sub-haplotype – The 5′ to 3′ sequence of nucleotides seen at a subset of the known polymorphic sites in a locus on a single chromosome from a single individual.

Haplotype pair – The two haplotypes found for a locus in a single individual.

Haplotyping – A process for determining one or more haplotypes in an individual and includes use of family pedigrees, molecular techniques and/or statistical inference.

Haplotype data - Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a population; frequency of each haplotype in that or other populations, and any known associations between one or more haplotypes and a trait.

Isoform – A particular form of a gene, mRNA, cDNA or the protein encoded thereby, distinguished from other forms by its particular sequence and/or structure.

Isogene – One of the isoforms of a gene found in a population. An isogene contains all of the polymorphisms present in the particular isoform of the gene.

Isolated – As applied to a biological molecule such as RNA, DNA, oligonucleotide, or protein, isolated means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally, the term "isolated" is not intended to refer to a complete absence of such material or to absence of water, buffers, or salts, unless they are present in amounts that substantially interfere with the methods of the present invention.

Locus - A location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature.

Naturally-occurring – A term used to designate that the object it is applied to, e.g., naturally-

occurring polynucleotide or polypeptide, can be isolated from a source in nature and which has not been intentionally modified by man.

Nucleotide pair – The nucleotides found at a polymorphic site on the two copies of a chromosome from an individual.

Phased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, phased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is known.

Polymorphic site (PS) – A position within a locus at which at least two alternative sequences are found in a population, the most frequent of which has a frequency of no more than 99%.

Polymorphic variant – A gene, mRNA, cDNA, polypeptide or peptide whose nucleotide or amino acid sequence varies from a reference sequence due to the presence of a polymorphism in the gene.

Polymorphism – The sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

Polymorphism data – Information concerning one or more of the following for a specific gene: location of polymorphic sites; sequence variation at those sites; frequency of polymorphisms in one or more populations; the different genotypes and/or haplotypes determined for the gene; frequency of one or more of these genotypes and/or haplotypes in one or more populations; any known association(s) between a trait and a genotype or a haplotype for the gene.

Polymorphism Database – A collection of polymorphism data arranged in a systematic or methodical way and capable of being individually accessed by electronic or other means.

Polynucleotide – A nucleic acid molecule comprised of single-stranded RNA or DNA or comprised of complementary, double-stranded DNA.

Population Group – A group of individuals sharing a common ethnogeographic origin.

Reference Population – A group of subjects or individuals who are predicted to be representative of the genetic variation found in the general population. Typically, the reference population represents the genetic variation in the population at a certainty level of at least 85%, preferably at least 90%, more preferably at least 95% and even more preferably at least 99%.

Single Nucleotide Polymorphism (SNP) – Typically, the specific pair of nucleotides observed at a single polymorphic site. In rare cases, three or four nucleotides may be found.

Subject – A human individual whose genotypes or haplotypes or response to treatment or disease state are to be determined.

Treatment - A stimulus administered internally or externally to a subject.

Unphased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, unphased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is not known.

The inventors herein have discovered 24 novel polymorphic sites, and confirmed the existence of

a 25th site, in the TNFRSF11B gene. The polymorphic sites identified by the inventors are referred to as PS1-25 to designate the order in which they are located in the gene (see Table 3 below), with the novel polymorphic sites referred to as PS1, PS2, PS3, PS4, PS5, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, and PS 25, and the previously reported polymorphic site referred to as PS6.

Thus, in one embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the TNFRSF11B gene or a fragment of the gene which contains at least one of the novel polymorphic sites described herein. The nucleotide sequence of a variant TNFRSF11B gene is identical to the reference genomic sequence for those portions of the gene examined, as described in the Examples below, except that it comprises a different nucleotide at one or more of the novel polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, and PS 25, and may also comprise an additional polymorphism of cytosine at PS6. Similarly, the nucleotide sequence of a variant fragment of the TNFRSF11B gene is identical to the corresponding portion of the reference sequence except for having a different nucleotide at one or more of the novel polymorphic sites described herein. Thus, the invention specifically does not include polynucleotides comprising a nucleotide sequence identical to the reference sequence (or other reported TNFRSF11B sequences) or to portions of the reference sequence (or other reported TNFRSF11B sequences), except for genotyping oligonucleotides as described below.

The location of a polymorphism in a variant gene or fragment is identified by aligning its sequence against SEQ ID NOS:1-2. The polymorphism is selected from the group consisting of adenine at PS1, thymine at PS2, thymine at PS3, thymine at PS4, cytosine at PS5, thymine at PS7, adenine at PS8, adenine at PS9, thymine at PS10, cytosine at PS11, cytosine at PS12, thymine at PS13, adenine at PS14, thymine at PS15, adenine at PS16, guanine at PS17, guanine at PS18, cytosine at PS19, adenine at PS20, cytosine at PS21, thymine at PS22, adenine at PS23, thymine at PS24 and cytosine at PS25. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the TNFRSF11B gene which is defined by any one of haplotypes 1-27 shown in Table 5 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the TNFRSF11B gene from a human genomic library. The clone may be sequenced to determine the identity of the nucleotides at the polymorphic sites described herein. Any particular variant claimed herein could be prepared from this clone by performing *in vitro* mutagenesis using procedures well-known in the art.

TNFRSF11B isogenes may be isolated using any method that allows separation of the two "copies" of the TNFRSF11B gene present in an individual, which, as readily understood by the skilled artisan, may be the same allele or different alleles. Separation methods include targeted *in vivo* cloning (TIVC) in yeast as described in WO 98/01573, U.S. Patent No. 5,866,404, and copending U.S. application Serial No. 08/987,966. Another method, which is described in copending U.S. Application Serial No. 08/987,966, uses an allele specific oligonucleotide in combination with primer extension and

exonuclease degradation to generate hemizygous DNA targets. Yet other methods are single molecule dilution (SMD) as described in Ruaño et al., Proc. Natl. Acad. Sci. 87:6296-6300, 1990; and allele specific PCR (Ruaño et al., 17 Nucleic Acids. Res. 8392, 1989; Ruaño et al., 19 Nucleic Acids Res. 6877-6882, 1991; Michalatos-Beloin et al., 24 Nucleic Acids Res. 4841-4843, 1996).

The invention also provides TNFRSF11B genome anthologies, which are collections of TNFRSF11B isogenes found in a given population. The population may be any group of at least two individuals, including but not limited to a reference population, a population group, a family population, a clinical population, and a same sex population. A TNFRSF11B genome anthology may comprise individual TNFRSF11B isogenes stored in separate containers such as microtest tubes, separate wells of a microtitre plate and the like. Alternatively, two or more groups of the TNFRSF11B isogenes in the anthology may be stored in separate containers. Individual isogenes or groups of isogenes in a genome anthology may be stored in any convenient and stable form, including but not limited to in buffered solutions, as DNA precipitates, freeze-dried preparations and the like. A preferred TNFRSF11B genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 5 below.

An isolated polynucleotide containing a polymorphic variant nucleotide sequence of the invention may be operably linked to one or more expression regulatory elements in a recombinant expression vector capable of being propagated and expressing the encoded TNFRSF11B protein in a prokaryotic or a eukaryotic host cell. Examples of expression regulatory elements which may be used include, but are not limited to, the lac system, operator and promoter regions of phage lambda, yeast promoters, and promoters derived from vaccinia virus, adenovirus, retroviruses, or SV40. Other regulatory elements include, but are not limited to, appropriate leader sequences, termination codons, polyadenylation signals, and other sequences required for the appropriate transcription and subsequent translation of the nucleic acid sequence in a given host cell. Of course, the correct combinations of expression regulatory elements will depend on the host system used. In addition, it is understood that the expression vector contains any additional elements necessary for its transfer to and subsequent replication in the host cell. Examples of such elements include, but are not limited to, origins of replication and selectable markers. Such expression vectors are commercially available or are readily constructed using methods known to those in the art (e.g., F. Ausubel et al., 1987, in "Current Protocols in Molecular Biology", John Wiley and Sons, New York, New York). Host cells which may be used to express the variant TNFRSF11B sequences of the invention include, but are not limited to, eukaryotic and mammalian cells, such as animal, plant, insect and yeast cells, and prokaryotic cells, such as E. coli, or algal cells as known in the art. The recombinant expression vector may be introduced into the host cell using any method known to those in the art including, but not limited to, microinjection, electroporation, particle bombardment, transduction, and transfection using DEAE-dextran, lipofection, or calcium phosphate (see e.g., Sambrook et al. (1989) in "Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In a preferred aspect, eukaryotic expression vectors that function in eukaryotic cells, and preferably mammalian cells, are used. Non-limiting examples of such vectors include vaccinia virus vectors,

adenovirus vectors, herpes virus vectors, and baculovirus transfer vectors. Preferred eukaryotic cell lines include COS cells, CHO cells, HeLa cells, NIH/3T3 cells, and embryonic stem cells (Thomson, J. A. et al., 1998 Science 282:1145-1147). Particularly preferred host cells are mammalian cells.

As will be readily recognized by the skilled artisan, expression of polymorphic variants of the TNFRSF11B gene will produce TNFRSF11B mRNAs varying from each other at any polymorphic site retained in the spliced and processed mRNA molecules. These mRNAs can be used for the preparation of a TNFRSF11B cDNA comprising a nucleotide sequence which is a polymorphic variant of the TNFRSF11B reference coding sequence shown in Figure 3. Thus, the invention also provides TNFRSF11B mRNAs and corresponding cDNAs which comprise a nucleotide sequence that is identical to SEQ ID NO:3 (Fig. 3), or its corresponding RNA sequence, except for having one or more polymorphisms selected from the group consisting of thymine at a position corresponding to nucleotide 699, adenine at a position corresponding to nucleotide 714, guanine at a position corresponding to nucleotide 720, guanine at a position corresponding to nucleotide 768, adenine at a position corresponding to nucleotide 841, thymine at a position corresponding to nucleotide 1102 and cytosine at a position corresponding to nucleotide 1150, and may also comprise an additional polymorphism of cytosine at a position corresponding to nucleotide 9 in Figure 3. Fragments of these variant mRNAs and cDNAs are included in the scope of the invention, provided they contain the novel polymorphisms described herein. The invention specifically excludes polynucleotides identical to previously identified and characterized TNFRSF11B cDNAs and fragments thereof. Polynucleotides comprising a variant RNA or DNA sequence may be isolated from a biological sample using well-known molecular biological procedures or may be chemically synthesized.

Genomic and cDNA fragments of the invention comprise at least one novel polymorphic site identified herein and have a length of at least 10 nucleotides and may range up to the full length of the gene. Preferably, a fragment according to the present invention is between 100 and 3000 nucleotides in length, and more preferably between 200 and 2000 nucleotides in length, and most preferably between 500 and 1000 nucleotides in length.

In describing the polymorphic sites identified herein, reference is made to the sense strand of the gene for convenience. However, as recognized by the skilled artisan, nucleic acid molecules containing the TNFRSF11B gene may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. Thus, reference may be made to the same polymorphic site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target region containing the polymorphic site. Thus, the invention also includes single-stranded polynucleotides which are complementary to the sense strand of the TNFRSF11B genomic variants described herein.

Polynucleotides comprising a polymorphic gene variant or fragment may be useful for therapeutic purposes. For example, where a patient could benefit from expression, or increased expression, of a particular TNFRSF11B protein isoform, an expression vector encoding the isoform may

be administered to the patient. The patient may be one who lacks the TNFRSF11B isogene encoding that isoform or may already have at least one copy of that isogene.

In other situations, it may be desirable to decrease or block expression of a particular TNFRSF11B isogene. Expression of a TNFRSF11B isogene may be turned off by transforming a targeted organ, tissue or cell population with an expression vector that expresses high levels of untranslatable mRNA for the isogene. Alternatively, oligonucleotides directed against the regulatory regions (e.g., promoter, introns, enhancers, 3′ untranslated region) of the isogene may block transcription. Oligonucleotides targeting the transcription initiation site, e.g., between positions –10 and +10 from the start site are preferred. Similarly, inhibition of transcription can be achieved using oligonucleotides that base-pair with region(s) of the isogene DNA to form triplex DNA (see e.g., Gee et al. in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y., 1994). Antisense oligonucleotides may also be designed to block translation of TNFRSF11B mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of TNFRSF11B mRNA transcribed from a particular isogene.

The oligonucleotides may be delivered to a target cell or tissue by expression from a vector introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, the oligonucleotides may be formulated as a pharmaceutical composition for administration to the patient. Oligoribonucleotides and/or oligodeoxynucleotides intended for use as antisense oligonucleotides may be modified to increase stability and half-life. Possible modifications include, but are not limited to phosphorothioate or 2′ O-methyl linkages, and the inclusion of nontraditional bases such as inosine and queosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous nucleases.

The invention also provides an isolated polypeptide comprising a polymorphic variant of the reference TNFRSF11B amino acid sequence shown in Figure 4 (SEQ ID NO:4). The location of a variant amino acid in a TNFRSF11B polypeptide or fragment of the invention is identified by aligning its sequence against Fig. 4. A TNFRSF11B protein variant of the invention comprises an amino acid sequence identical to SEQ ID NO: 4 except for having one or more variant amino acids selected from the group consisting of methionine at a position corresponding to amino acid position 240, methionine at a position corresponding to amino acid position 281, and serine at a position corresponding to amino acid 368, and may also comprise an additional variant amino acid of asparagine at a position corresponding to amino acid 3 in Figure 4. The invention specifically excludes amino acid sequences identical to those previously identified for TNFRSF11B, including SEQ ID NO: 4, and previously described fragments thereof. TNFRSF11B protein variants included within the invention comprise all amino acid sequences based on SEQ ID NO: 4 and having the combination of amino acid variations described in Table 2 below. In preferred embodiments, a TNFRSF11B protein variant of the invention is encoded by an isogene defined by one of the observed haplotypes shown in Table 5.

Table 2. Novel Protein Variants of TNFRSF11B				
Isoform No.	Amino Acid Positions			
	3	240	281	368
1	N	M	M	S
2	N	M	M	T
3	N	M	V	S
4	N	M	V	T
5	N	I	M	S
6	N	I	M	T
7	N	I	V	S
8	K	M	M	S
9	K	M	M	T
10	K	M	V	S
11	K	M	V	T
12	K	I	M	S
13	K	I	M	T
14	K	I	V	S

The invention also includes TNFRSF11B peptide variants, which are any fragments of a TNFRSF11B protein variant that contains one or more of the novel amino acid variations shown in Table 2. A TNFRSF11B peptide variant is at least 6 amino acids in length and is preferably any number between 6 and 30 amino acids long, more preferably between 10 and 25, and most preferably between 15 and 20 amino acids long. Such TNFRSF11B peptide variants may be useful as antigens to generate antibodies specific for one of the above TNFRSF11B isoforms. In addition, the TNFRSF11B peptide variants may be useful in drug screening assays.

A TNFRSF11B variant protein or peptide of the invention may be prepared by chemical synthesis or by expressing one of the variant TNFRSF11B genomic and cDNA sequences as described above. Alternatively, the TNFRSF11B protein variant may be isolated from a biological sample of an individual having a TNFRSF11B isogene which encodes the variant protein. Where the sample contains two different TNFRSF11B isoforms (i.e., the individual has different TNFRSF11B isogenes), a particular TNFRSF11B isoform of the invention can be isolated by immunoaffinity chromatography using an antibody which specifically binds to that particular TNFRSF11B isoform but does not bind to the other TNFRSF11B isoform.

The expressed or isolated TNFRSF11B protein may be detected by methods known in the art, including Coomassie blue staining, silver staining, and Western blot analysis using antibodies specific for the isoform of the TNFRSF11B protein as discussed further below. TNFRSF11B variant proteins can be purified by standard protein purification procedures known in the art, including differential precipitation, molecular sieve chromatography, ion-exchange chromatography, isoelectric focusing, gel electrophoresis, affinity and immunoaffinity chromatography and the like. (Ausubel et. al., 1987, In Current Protocols in Molecular Biology John Wiley and Sons, New York, New York). In the case of immunoaffinity chromatography, antibodies specific for a particular polymorphic variant may be used.

A polymorphic variant TNFRSF11B gene of the invention may also be fused in frame with a heterologous sequence to encode a chimeric TNFRSF11B protein. The non-TNFRSF11B portion of the chimeric protein may be recognized by a commercially available antibody. In addition, the chimeric protein may also be engineered to contain a cleavage site located between the TNFRSF11B and non-TNFRSF11B portions so that the TNFRSF11B protein may be cleaved and purified away from the non-TNFRSF11B portion.

An additional embodiment of the invention relates to using a novel TNFRSF11B protein isoform in any of a variety of drug screening assays. Such screening assays may be performed to identify agents that bind specifically to all known TNFRSF11B protein isoforms or to only a subset of one or more of these isoforms. The agents may be from chemical compound libraries, peptide libraries and the like. The TNFRSF11B protein or peptide variant may be free in solution or affixed to a solid support. In one embodiment, high throughput screening of compounds for binding to a TNFRSF11B variant may be accomplished using the method described in PCT application WO84/03565, in which large numbers of test compounds are synthesized on a solid substrate, such as plastic pins or some other surface, contacted with the TNFRSF11B protein(s) of interest and then washed. Bound TNFRSF11B protein(s) are then detected using methods well-known in the art.

In another embodiment, a novel TNFRSF11B protein isoform may be used in assays to measure the binding affinities of one or more candidate drugs targeting the TNFRSF11B protein.

In another embodiment, the invention provides antibodies specific for and immunoreactive with one or more of the novel TNFRSF11B variant proteins described herein. The antibodies may be either monoclonal or polyclonal in origin. The TNFRSF11B protein or peptide variant used to generate the antibodies may be from natural or recombinant sources or produced by chemical synthesis using synthesis techniques known in the art. If the TNFRSF11B protein variant is of insufficient size to be antigenic, it may be conjugated, complexed, or otherwise covalently linked to a carrier molecule to enhance the antigenicity of the peptide. Examples of carrier molecules, include, but are not limited to, albumins (e.g., human, bovine, fish, ovine), and keyhole limpet hemocyanin (Basic and Clinical Immunology, 1991, Eds. D.P. Stites, and A.I. Terr, Appleton and Lange, Norwalk Connecticut, San Mateo, California).

In one embodiment, an antibody specifically immunoreactive with one of the novel TNFRSF11B protein isoforms described herein is administered to an individual to neutralize activity of the TNFRSF11B isoform expressed by that individual. The antibody may be formulated as a pharmaceutical composition which includes a pharmaceutically acceptable carrier.

Antibodies specific for and immunoreactive with one of the novel TNFRSF11B protein isoform described herein may be used to immunoprecipitate the TNFRSF11B protein variant from solution as well as react with TNFRSF11B protein isoforms on Western or immunoblots of polyacrylamide gels on membrane supports or substrates. In another preferred embodiment, the antibodies will detect TNFRSF11B protein isoforms in paraffin or frozen tissue sections, or in cells which have been fixed or unfixed and prepared on slides, coverslips, or the like, for use in immunocytochemical,

immunohistochemical, and immunofluorescence techniques.

In another embodiment, an antibody specifically immunoreactive with one of the novel TNFRSF11B protein variants described herein is used in immunoassays to detect this variant in biological samples. In this method, an antibody of the present invention is contacted with a biological sample and the formation of a complex between the TNFRSF11B protein variant and the antibody is detected. As described, suitable immunoassays include radioimmunoassay, Western blot assay, immunofluorescent assay, enzyme linked immunoassay (ELISA), chemiluminescent assay, immunohistochemical assay, immunocytochemical assay, and the like (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Press, New York, New York; Current Protocols in Molecular Biology, 1987, Eds. Ausubel et al., John Wiley and Sons, New York, New York). Standard techniques known in the art for ELISA are described in Methods in Immunodiagnosis, 2nd Ed., Eds. Rose and Bigazzi, John Wiley and Sons, New York 1980; and Campbell et al., 1984, Methods in Immunology, W.A. Benjamin, Inc.). Such assays may be direct, indirect, competitive, or noncompetitive as described in the art (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Pres, NY, NY; and Oellirich, M., 1984, J. Clin. Chem. Clin. Biochem., 22:895-904). Proteins may be isolated from test specimens and biological samples by conventional methods, as described in Current Protocols in Molecular Biology, supra.

Exemplary antibody molecules for use in the detection and therapy methods of the present invention are intact immunoglobulin molecules, substantially intact immunoglobulin molecules, or those portions of immunoglobulin molecules that contain the antigen binding site. Polyclonal or monoclonal antibodies may be produced by methods conventionally known in the art (e.g., Kohler and Milstein, 1975, Nature, 256:495-497; Campbell Monoclonal Antibody Technology, the Production and Characterization of Rodent and Human Hybridomas, 1985, In: Laboratory Techniques in Biochemistry and Molecular Biology, Eds. Burdon et al., Volume 13, Elsevier Science Publishers, Amsterdam). The antibodies or antigen binding fragments thereof may also be produced by genetic engineering. The technology for expression of both heavy and light chain genes in E. coli is the subject of PCT patent applications, publication number WO 901443, WO 901443 and WO 9014424 and in Huse et al., 1989, Science, 246:1275-1281. The antibodies may also be humanized (e.g., Queen, C. et al. 1989 Proc. Natl. Acad. Sci. 86;10029).

Effect(s) of the polymorphisms identified herein on expression of TNFRSF11B may be investigated by preparing recombinant cells and/or organisms, preferably recombinant animals, containing a polymorphic variant of the TNFRSF11B gene. As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA into TNFRSF11B protein (including codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

To prepare a recombinant cell of the invention, the desired TNFRSF11B isogene may be introduced into the cell in a vector such that the isogene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the TNFRSF11B isogene is introduced into a cell in such a way that it recombines with the endogenous TNFRSF11B gene present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired TNFRSF11B gene polymorphism. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner. Examples of cells into which the TNFRSF11B isogene may be introduced include, but are not limited to, continuous culture cells, such as COS, NIH/3T3, and primary or culture cells of the relevant tissue type, i.e., they express the TNFRSF11B isogene. Such recombinant cells can be used to compare the biological activities of the different protein variants.

Recombinant organisms, i.e., transgenic animals, expressing a variant TNFRSF11B gene are prepared using standard procedures known in the art. Preferably, a construct comprising the variant gene is introduced into a nonhuman animal or an ancestor of the animal at an embryonic stage, i.e., the one-cell stage, or generally not later than about the eight-cell stage. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill in the art. One method involves transfecting into the embryo a retrovirus constructed to contain one or more insulator elements, a gene or genes of interest, and other components known to those skilled in the art to provide a complete shuttle vector harboring the insulated gene(s) as a transgene, see e.g., U.S. Patent No. 5,610,053. Another method involves directly injecting a transgene into the embryo. A third method involves the use of embryonic stem cells. Examples of animals into which the TNFRSF11B isogenes may be introduced include, but are not limited to, mice, rats, other rodents, and nonhuman primates (see "The Introduction of Foreign Genes into Mice" and the cited references therein, In: Recombinant DNA, Eds. J.D. Watson, M. Gilman, J. Witkowski, and M. Zoller; W.H. Freeman and Company, New York, pages 254-272). Transgenic animals stably expressing a human TNFRSF11B isogene and producing human TNFRSF11B protein can be used as biological models for studying diseases related to abnormal TNFRSF11B expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

An additional embodiment of the invention relates to pharmaceutical compositions for treating disorders affected by expression or function of a novel TNFRSF11B isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel TNFRSF11B isogenes; an antisense oligonucleotide directed against one of the novel TNFRSF11B isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another

compound which inhibits expression of a novel TNFRSF11B isogene described herein. Preferably, the composition contains the active ingredient in a therapeutically effective amount. By therapeutically effective amount is meant that one or more of the symptoms relating to disorders affected by expression or function of a novel TNFRSF11B isogene is reduced and/or eliminated. The composition also comprises a pharmaceutically acceptable carrier, examples of which include, but are not limited to, saline, buffered saline, dextrose, and water. Those skilled in the art may employ a formulation most suitable for the active ingredient, whether it is a polynucleotide, oligonucleotide, protein, peptide or small molecule antagonist. The pharmaceutical composition may be administered alone or in combination with at least one other agent, such as a stabilizing compound. Administration of the pharmaceutical composition may be by any number of routes including, but not limited to oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, intradermal, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

For any composition, determination of the therapeutically effective dose of active ingredient and/or the appropriate route of administration is well within the capability of those skilled in the art. For example, the dose can be estimated initially either in cell culture assays or in animal models. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. The exact dosage will be determined by the practitioner, in light of factors relating to the patient requiring treatment, including but not limited to severity of the disease state, general health, age, weight and gender of the patient, diet, time and frequency of administration, other drugs being taken by the patient, and tolerance/response to the treatment.

Information on the identity of genotypes and haplotypes for the TNFRSF11B gene of any particular individual as well as the frequency of such genotypes and haplotypes in any particular population of individuals is expected to be useful for a variety of basic research and clinical applications. Thus, the invention also provides compositions and methods for detecting the novel TNFRSF11B polymorphisms identified herein.

The compositions comprise at least one TNFRSF11B genotyping oligonucleotide. In one embodiment, a TNFRSF11B genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that is located close to, or that contains, one of the novel polymorphic sites described herein. As used herein, the term "oligonucleotide" refers to a polynucleotide molecule having less than about 100 nucleotides. A preferred oligonucleotide of the invention is 10 to 35 nucleotides long. More preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The oligonucleotide may be comprised of any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives, and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of

linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995), pages 617-620). Oligonucleotides of the invention may be prepared by chemical synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the art, including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags and the like.

Genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of a TNFRSF11B polynucleotide, i.e., a TNFRSF11B isogene. As used herein, specific hybridization means the oligonucleotide forms an anti-parallel double-stranded structure with the target region under certain hybridizing conditions, while failing to form such a structure when incubated with a non-target region or a non-TNFRSF11B polynucleotide under the same hybridizing conditions. Preferably, the oligonucleotide specifically hybridizes to the target region under conventional high stringency conditions. The skilled artisan can readily design and test oligonucleotide probes and primers suitable for detecting polymorphisms in the TNFRSF11B gene using the polymorphism information provided herein in conjunction with the known sequence information for the TNFRSF11B gene and routine techniques.

A nucleic acid molecule such as an oligonucleotide or polynucleotide is said to be a "perfect" or "complete" complement of another nucleic acid molecule if every nucleotide of one of the molecules is complementary to the nucleotide at the corresponding position of the other molecule. A nucleic acid molecule is "substantially complementary" to another molecule if it hybridizes to that molecule with sufficient stability to remain in a duplex form under conventional low-stringency conditions. Conventional hybridization conditions are described, for example, by Sambrook J. et al., in Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989) and by Haymes, B.D. et al. in Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, D.C. (1985). While perfectly complementary oligonucleotides are preferred for detecting polymorphisms, departures from complete complementarity are contemplated where such departures do not prevent the molecule from specifically hybridizing to the target region. For example, an oligonucleotide primer may have a non-complementary fragment at its 5′ end, with the remainder of the primer being complementary to the target region. Alternatively, non-complementary nucleotides may be interspersed into the oligonucleotide probe or primer as long as the resulting probe or primer is still capable of specifically hybridizing to the target region.

Preferred genotyping oligonucleotides of the invention are allele-specific oligonucleotides. As used herein, the term allele-specific oligonucleotide (ASO) means an oligonucleotide that is able, under sufficiently stringent conditions, to hybridize specifically to one allele of a gene, or other locus, at a target region containing a polymorphic site while not hybridizing to the corresponding region in another allele(s). As understood by the skilled artisan, allele-specificity will depend upon a variety of readily

optimized stringency conditions, including salt and formamide concentrations, as well as temperatures for both the hybridization and washing steps. Examples of hybridization and washing conditions typically used for ASO probes are found in Kogan et al., "Genetic Prediction of Hemophilia A" in PCR Protocols, A Guide to Methods and Applications, Academic Press, 1990 and Ruano et al., 87 Proc. Natl. Acad. Sci. USA 6296-6300, 1990. Typically, an allele-specific oligonucleotide will be perfectly complementary to one allele while containing a single mismatch for another allele.

Allele-specific oligonucleotide probes which usually provide good discrimination between different alleles are those in which a central position of the oligonucleotide probe aligns with the polymorphic site in the target region (e.g., approximately the 7th or 8th position in a 15 mer, the 8th or 9th position in a 16mer, the 10th or 11th position in a 20 mer). A preferred ASO probe for detecting TNFRSF11B gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

Accession No.:AB008821
ACTTGAAGATGAATG (SEQ ID NO:5) and its complement,
ACTTGAAAATGAATG (SEQ ID NO: 6) and its complement,

GATCTTGGCTGGATC (SEQ ID NO: 7) and its complement,
GATCTTGTCTGGATC (SEQ ID NO: 8) and its complement,

CCACCGCCCCACCCC (SEQ ID NO: 9) and its complement,
CCACCGCTCCACCCC (SEQ ID NO: 10) and its complement,

TCCCTGGGGGATCCT (SEQ ID NO: 11) and its complement,
TCCCTGGTGGATCCT (SEQ ID NO: 12) and its complement,

GCGTTAATCCTGGAG (SEQ ID NO: 13) and its complement,
GCGTTAACCCTGGAG (SEQ ID NO: 14) and its complement,

CCTGGGCCAGCCGAC (SEQ ID NO: 15) and its complement,
CCTGGGCTAGCCGAC (SEQ ID NO: 16) and its complement,

GGGAGAAGGCTCCAC (SEQ ID NO: 17) and its complement,
GGGAGAAAGCTCCAC (SEQ ID NO: 18) and its complement,

CCTTTTACGCTGCAA (SEQ ID NO: 19) and its complement,
CCTTTTAAGCTGCAA (SEQ ID NO: 20) and its complement,

Accession No.:AB008822
GCTGGTACGTGTCAA (SEQ ID NO: 21) and its complement,
GCTGGTATGTGTCAA (SEQ ID NO: 22) and its complement,

AGGACCATTGCTCAG (SEQ ID NO: 23) and its complement,
AGGACCACTGCTCAG (SEQ ID NO: 24) and its complement,

AACATAATAGTAGCA (SEQ ID NO: 25) and its complement,
AACATAACAGTAGCA (SEQ ID NO: 26) and its complement,

TATTTTCCGTAGGAA (SEQ ID NO: 27) and its complement,
TATTTTCTGTAGGAA (SEQ ID NO: 28) and its complement,

CATTTTAGCATATTT (SEQ ID NO: 29) and its complement,
CATTTTAACATATTT (SEQ ID NO: 30) and its complement,

AAGTAAACGCAGAGA (SEQ ID NO: 31) and its complement,
AAGTAAATGCAGAGA (SEQ ID NO: 32) and its complement,

GTGTAGAGAGGATAA (SEQ ID NO: 33) and its complement,
GTGTAGAAAGGATAA (SEQ ID NO: 34) and its complement,

AGAGGATAAAACGGC (SEQ ID NO: 35) and its complement,
AGAGGATGAAACGGC (SEQ ID NO: 36) and its complement,

TGAAGTTATGGAAAC (SEQ ID NO: 37) and its complement,
TGAAGTTGTGGAAAC (SEQ ID NO: 38) and its complement,

TGAAGTTATGGAAAC (SEQ ID NO: 39) and its complement,
TGAAGTTGTGGAAAC (SEQ ID NO: 40) and its complement,

GTATGATAATCTAAA (SEQ ID NO: 41) and its complement,
GTATGATCATCTAAA (SEQ ID NO: 42) and its complement,

TAGTTACGGCAATTA SEQ ID NO: 43) and its complement;
TAGTTACAGCAATTA SEQ ID NO: 44) and its complement,

AGAGTGATGTGTCTT SEQ ID NO: 45) and its complement,
AGAGTGACGTGTCTT SEQ ID NO: 46) and its complement,

GGCATTGGATGAATAT SEQ ID NO: 47) and its complement,
GGCATTGTATGAATAT SEQ ID NO: 48) and its complement,

AAACAGCGTGCAGCG (SEQ ID NO: 49) and its complement,
AAACAGCATGCAGCG (SEQ ID NO: 50) and its complement,

AAAGAAGACCATCAG SEQ ID NO: 51) and its complement,
AAAGAAGTCCATCAG SEQ ID NO: 52) and its complement,

TCAGAAGTTATTTTT (SEQ ID NO: 53) and its complement, and
TCAGAAGCTATTTTT (SEQ ID NO: 54) and its complement.

An allele-specific oligonucleotide primer of the invention has a 3′ terminal nucleotide, or preferably a 3′ penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. Allele-specific oligonucleotide primers hybridizing to either the coding or noncoding strand are contemplated by the invention. A preferred ASO primer for detecting TNFRSF11B gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

Accession No.:AB008821
ATGTAAACTTGAAGA (SEQ ID NO: 55); TTCGCAATCATTCATCT (SEQ ID NO: 56)
ATGTAAACTTGAAAA (SEQ ID NO: 57); TTCGCAATCATTCATTT (SEQ ID NO: 58)

CGTCCGGATCTTGGC (SEQ ID NO: 59); GAGTCCGATCCAGCC (SEQ ID NO: 60);
CGTCCGGATCTTGTC (SEQ ID NO: 61); GAGTCCGATCCAGAC (SEQ ID NO: 62);

CAGACACCACCGCCC (SEQ ID NO: 63); GCGTGAGGGGTGGGG (SEQ ID NO: 64);
CAGACACCACCGCTC (SEQ ID NO: 65); GCGTGAGGGGTGGAG (SEQ ID NO: 66);

CCCACCTCCCTGGGG (SEQ ID NO: 67); GCGGAAAGGATCCCC (SEQ ID NO: 68);
CCCACCTCCCTGGTG (SEQ ID NO: 69); GCGGAAAGGATCCAC (SEQ ID NO: 70);

CTGAAAGCGTTAATC (SEQ ID NO: 71); AGAAAGCTCCAGGAT (SEQ ID NO: 72);
CTGAAAGCGTTAACC (SEQ ID NO: 73); AGAAAGCTCCAGGGT (SEQ ID NO: 74);

TAAGTCCCTGGGCCA (SEQ ID NO: 75); GCACCCGTCGGCTGG (SEQ ID NO: 76);
TAAGTCCCTGGGCTA (SEQ ID NO: 77); GCACCCGTCGGCTAG (SEQ ID NO: 78);

CCGGCGGGGAGAAGG (SEQ ID NO: 79); GAGCGAGTGGAGCCT (SEQ ID NO: 80);
CCGGCGGGGAGAAAG (SEQ ID NO: 81); GAGCGAGTGGAGCTT (SEQ ID NO: 82);

GGGTGTCCTTTTACG (SEQ ID NO: 83); GGAACTTTGCAGCGT (SEQ ID NO: 84);
GGGTGTCCTTTTAAG (SEQ ID NO: 85); GGAACTTTGCAGCTT (SEQ ID NO: 86);

Accession No.:AB008822
GTGCAAGCTGGTACG (SEQ ID NO: 87); TGCACATTGACACGT (SEQ ID NO: 88);
GTGCAAGCTGGTATG (SEQ ID NO: 89); TGCACATTGACACAT (SEQ ID NO: 90);

TTCCAAAGGACCATT (SEQ ID NO: 91); ATTCCTCTGAGCAAT (SEQ ID NO: 92);
TTCCAAAGGACCACT (SEQ ID NO: 93); ATTCCTCTGAGCAGT (SEQ ID NO: 94);

TTGTGCAACATAATA (SEQ ID NO: 95); TTTTACTGCTACTAT (SEQ ID NO: 96);
TTGTGCAACATAACA (SEQ ID NO: 97); TTTTACTGCTACTGT (SEQ ID NO: 98);

GCTTGGTATTTTCCG (SEQ ID NO: 99); CTGGGGTTCCTACGG (SEQ ID NO: 100);
GCTTGGTATTTTCTG (SEQ ID NO: 101); CTGGGGTTCCTACAG (SEQ ID NO: 102);

ACTTTGCATTTTAGC (SEQ ID NO: 103); AAGATAAAATATGCT (SEQ ID NO: 104);
ACTTTGCATTTTAAC (SEQ ID NO: 105); AAGATAAAATATGTT (SEQ ID NO: 106);

GCACCAAAGTAAACG (SEQ ID NO: 107); CTACACTCTCTGCGT (SEQ ID NO: 108);
GCACCAAAGTAAATG (SEQ ID NO: 109); CTACACTCTCTGCAT (SEQ ID NO: 110);

CAGAGAGTGTAGAGA (SEQ ID NO: 111) CGCGTTTTATCCT (SEQ ID NO: 112)
CAGAGAGTGTAGAAA (SEQ ID NO: 113) CGCGTTTTATCTT (SEQ ID NO: 114)

GTGTAGAGAGGATAA (SEQ ID NO: 115); TGTGTTGCCGTTTTA (SEQ ID NO: 116);
GTGTAGAGAGGATGA (SEQ ID NO: 117); TGTGTTGCCGTTTCA (SEQ ID NO: 118);

AGCTGCTGAAGTTAT (SEQ ID NO: 119); TTTGATGTTTCCATA (SEQ ID NO: 120);
AGCTGCTGAAGTTGT (SEQ ID NO: 121); TTTGATGTTTCCACA (SEQ ID NO: 122);

TCCAAGGTATGATAA (SEQ ID NO: 123); TTTTATTTTAGATTA (SEQ ID NO: 124);
TCCAAGGTATGATCA (SEQ ID NO: 125); TTTTATTTTAGATGA (SEQ ID NO: 126);

AAAGGCTAGTTACGG (SEQ ID NO: 127); TGATAAGTTAATTGC (SEQ ID NO: 128)
AAAGGCTAGTTACAG (SEQ ID NO: 129); TGATAAGTTAATTTC (SEQ ID NO: 130)

AAACTGAGAGTGATG (SEQ ID NO: 131); GAAAATAAGACACAT (SEQ ID NO: 132)
AAACTGAGAGTGACG (SEQ ID NO: 133); GAAAATAAGACACGT (SEQ ID NO: 134)

TATGATGGCATTGGA (SEQ ID NO: 135); CATTTATATTCATCC (SEQ ID NO: 136)
TATGATGGCATTGTA (SEQ ID NO: 137); CATTTATATTCATAC (SEQ ID NO: 138)

CTGTGAAAACAGCGT (SEQ ID NO: 139); ATGTGCCGCTGCACG (SEQ ID NO: 140);

CTGTGAAAACAGCAT (SEQ ID NO: 141); ATGTGCCGCTGCATG (SEQ ID NO: 142);

GAGTCTAAAGAAGAC (SEQ ID NO: 143); AGGAACCTGATGGTC (SEQ ID NO: 144);
GAGTCTAAAGAAGTC (SEQ ID NO: 145); AGGAACCTGATGGAC (SEQ ID NO: 146);

ATTGTATCAGAAGTT (SEQ ID NO: 147); ATTTCTAAAAATAAC (SEQ ID NO: 148);
ATTGTATCAGAAGCT (SEQ ID NO: 149); and ATTTCTAAAAATAGC (SEQ ID NO:150).

Other genotyping oligonucleotides of the invention hybridize to a target region located one to several nucleotides downstream of one of the novel polymorphic sites identified herein. Such oligonucleotides are useful in polymerase-mediated primer extension methods for detecting one of the novel polymorphisms described herein and therefore such genotyping oligonucleotides are referred to herein as "primer-extension oligonucleotides". In a preferred embodiment, the 3′-terminus of a primer-extension oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately adjacent to the polymorphic site. A particularly preferred oligonucleotide primer for detecting TNFRSF11B gene polymorphisms by primer extension terminates in a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

Accession No.:AB008821

TAAACTTGAA	(SEQ ID NO: 151);	AATCATTCAT	(SEQ ID NO: 152);
CCGGATCTTG	(SEQ ID NO: 153);	TCCGATCCAG	(SEQ ID NO: 154);
ACACCACCGC	(SEQ ID NO: 155);	TGAGGGGTGG	(SEQ ID NO: 156);
ACCTCCCTGG	(SEQ ID NO: 157);	GAAAGGATCC	(SEQ ID NO: 158);
AAAGCGTTAA	(SEQ ID NO: 159);	AAGCTCCAGG	(SEQ ID NO: 160);
GTCCCTGGGC	(SEQ ID NO: 161);	CCCGTCGGCT	(SEQ ID NO: 162);
GCGGGGAGAA	(SEQ ID NO: 163);	CGAGTGGAGC	(SEQ ID NO: 164);
TGTCCTTTTA	(SEQ ID NO: 165);	ACTTTGCAGC	(SEQ ID NO: 166);

Accession No.:AB008822

CAAGCTGGTA	(SEQ ID NO: 167);	ACATTGACAC	(SEQ ID NO: 168);
CAAAGGACCA	(SEQ ID NO: 169);	CCTCTGAGCA	(SEQ ID NO: 170);
TGCAACATAA	(SEQ ID NO: 171);	TACTGCTACT	(SEQ ID NO: 172);
TGGTATTTTC	(SEQ ID NO: 173);	GGGTTCCTAC	(SEQ ID NO: 174);
TTGCATTTTA	(SEQ ID NO: 175);	ATAAAATATG	(SEQ ID NO: 176);
CCAAAGTAAA	(SEQ ID NO: 177);	CACTCTCTGC	(SEQ ID NO: 178);
AGAGTGTAGA	(SEQ ID NO: 179);	GTTTTATCCT	(SEQ ID NO: 180)
TAGAGAGGAT	(SEQ ID NO: 181);	GTTGCCGTTT	(SEQ ID NO: 182);
TGCTGAAGTT	(SEQ ID NO: 183);	GATGTTTCCA	(SEQ ID NO: 184);
AAGGTATGAT	(SEQ ID NO: 185);	TATTTTAGAT	(SEQ ID NO: 186);
GGCTAGTTAC	(SEQ ID NO: 187);	AGTTAATTGC	(SEQ ID NO: 188);
CTGAGAGTGA	(SEQ ID NO: 189);	AATAAGACAC	(SEQ ID NO: 190);
GATGGCATTG	(SEQ ID NO: 191);	TTATATTCAT	(SEQ ID NO: 192);
TGAAAACAGC	(SEQ ID NO: 193);	TGCCGCTGCA	(SEQ ID NO: 194);
TCTAAAGAAG	(SEQ ID NO: 195);	AACCTGATGG	(SEQ ID NO: 196);
GTATCAGAAG	(SEQ ID NO: 197);	and TCTAAAAATA	(SEQ ID NO: 198).

In some embodiments, a composition contains two or more differently labeled genotyping oligonucleotides for simultaneously probing the identity of nucleotides at two or more polymorphic sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a polymorphic

site.

TNFRSF11B genotyping oligonucleotides of the invention may also be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, e.g., WO 98/20020 and WO 98/20019). Such immobilized genotyping oligonucleotides may be used in a variety of polymorphism detection assays, including but not limited to probe hybridization and polymerase extension assays. Immobilized TNFRSF11B genotyping oligonucleotides of the invention may comprise an ordered array of oligonucleotides designed to rapidly screen a DNA sample for polymorphisms in multiple genes at the same time.

In another embodiment, the invention provides a kit comprising at least two genotyping oligonucleotides packaged in separate containers. The kit may also contain other components such as hybridization buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as PCR.

The above described oligonucleotide compositions and kits are useful in methods for genotyping and/or haplotyping the TNFRSF11B gene in an individual. As used herein, the terms "TNFRSF11B genotype" and "TNFRSF11B haplotype" mean the genotype or haplotype contains the nucleotide pair or nucleotide, respectively, that is present at one or more of the novel polymorphic sites described herein and may optionally also include the nucleotide pair or nucleotide present at one or more additional polymorphic sites in the TNFRSF11B gene. The additional polymorphic sites may be currently known polymorphic sites or sites that are subsequently discovered.

One embodiment of the genotyping method involves isolating from the individual a nucleic acid mixture comprising the two copies of the TNFRSF11B gene, or a fragment thereof, that are present in the individual, and determining the identity of the nucleotide pair at one or more of the polymorphic sites selected from PS1, PS2, PS3, PS4, PS5, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, and PS 25 in the two copies to assign a TNFRSF11B genotype to the individual. As will be readily understood by the skilled artisan, the two "copies" of a gene in an individual may be the same allele or may be different alleles. In a preferred embodiment of the genotyping method, the identity of the nucleotide pair at PS6 is also determined. In a particularly preferred embodiment, the genotyping method comprises determining the identity of the nucleotide pair at each of PS1-25.

Typically, the nucleic acid mixture is isolated from a biological sample taken from the individual, such as a blood sample or tissue sample. Suitable tissue samples include whole blood, semen saliva, tears, urine, fecal material, sweat, buccal, skin and hair. The nucleic acid mixture may be comprised of genomic DNA, mRNA, or cDNA and, in the latter two cases, the biological sample must be obtained from an organ in which the TNFRSF11B gene is expressed. Furthermore it will be understood by the skilled artisan that mRNA or cDNA preparations would not be used to detect polymorphisms located in

introns or in 5′ and 3′ nontranscribed regions. If a TNFRSF11B gene fragment is isolated, it must contain the polymorphic site(s) to be genotyped.

One embodiment of the haplotyping method comprises isolating from the individual a nucleic acid molecule containing only one of the two copies of the TNFRSF11B gene, or a fragment thereof, that is present in the individual and determining in that copy the identity of the nucleotide at one or more of the polymorphic sites PS1, PS2, PS3, PS4, PS5, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, and PS 25 in that copy to assign a TNFRSF11B haplotype to the individual. The nucleic acid may be isolated using any method capable of separating the two copies of the TNFRSF11B gene or fragment such as one of the methods described above for preparing TNFRSF11B isogenes, with targeted *in vivo* cloning being the preferred approach. As will be readily appreciated by those skilled in the art, any individual clone will only provide haplotype information on one of the two TNFRSF11B gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional TNFRSF11B clones will need to be examined. Typically, at least five clones should be examined to have more than a 90% probability of haplotyping both copies of the TNFRSF11B gene in an individual. In some embodiments, the haplotyping method also comprises identifying the nucleotide at PS6. In a particularly preferred embodiment, the nucleotide at each of PS1-25 is identified.

In a preferred embodiment, a TNFRSF11B haplotype pair is determined for an individual by identifying the phased sequence of nucleotides at one or more of the polymorphic sites selected from PS1, PS2, PS3, PS4, PS5, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, and PS 25 in each copy of the TNFRSF11B gene that is present in the individual. In a particularly preferred embodiment, the haplotyping method comprises identifying the phased sequence of nucleotides at each of PS1-25 in each copy of the TNFRSF11B gene. When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, it is also envisioned that if the two copies are labeled with different tags, or are otherwise separately distinguishable or identifiable, it could be possible in some cases to perform the method in the same container. For example, if first and second copies of the gene are labeled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labeled with yet a third different fluorescent dye is used to assay the polymorphic site(s), then detecting a combination of the first and third dyes would identify the polymorphism in the first gene copy while detecting a combination of the second and third dyes would identify the polymorphism in the second gene copy.

In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic site(s) may be determined by amplifying a target region(s) containing the polymorphic site(s) directly from one or both copies of the TNFRSF11B gene, or fragment thereof, and the sequence of the amplified region(s) determined by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic site in individuals who are homozygous

at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example, where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for an individual homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

In addition, the identity of the allele(s) present at any of the novel polymorphic sites described herein may be indirectly determined by genotyping a polymorphic site not disclosed herein that is in linkage disequilibrium with the polymorphic site that is of interest. Two sites are said to be in linkage disequilibrium if the presence of a particular variant at one site enhances the predictability of another variant at the second site (Stevens, JC 1999, *Mol. Diag.* 4: 309-17). Polymorphic sites in linkage disequilibrium with the presently disclosed polymorphic sites may be located in regions of the gene or in other genomic regions not examined herein. Genotyping of a polymorphic site in linkage disequilibrium with the novel polymorphic sites described herein may be performed by, but is not limited to, any of the above-mentioned methods for detecting the identity of the allele at a polymorphic site.

The target region(s) may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR) (U.S. Patent No. 4,965,188), ligase chain reaction (LCR) (Barany et al., Proc. Natl. Acad. Sci. USA 88:189-193, 1991; WO90/01069), and oligonucleotide ligation assay (OLA) (Landegren et al., Science 241:1077-1080, 1988). Oligonucleotides useful as primers or probes in such methods should specifically hybridize to a region of the nucleic acid that contains or is adjacent to the polymorphic site. Typically, the oligonucleotides are between 10 and 35 nucleotides in length and preferably, between 15 and 30 nucleotides in length. Most preferably, the oligonucleotides are 20 to 25 nucleotides long. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan.

Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Patent No. 5,130,238; EP 329,822; U.S. Patent No. 5,169,766, WO89/06700) and isothermal methods (Walker et al., *Proc. Natl. Acad. Sci. USA* 89:392-396, 1992).

A polymorphism in the target region may also be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically, allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labeled probe pairs, with one member of the pair showing a perfect match to one variant of a target sequence and the other member showing a perfect match to a different variant. In some embodiments, more than one polymorphic site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of the polymorphic sites being

detected.

Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Allele-specific oligonucleotides may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibers, chips, dishes, and beads. The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

The genotype or haplotype for the TNFRSF11B gene of an individual may also be determined by hybridization of a nucleic sample containing one or both copies of the gene to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would contain a battery of allele-specific oligonucleotides representing each of the polymorphic sites to be included in the genotype or haplotype.

The identity of polymorphisms may also be determined using a mismatch detection technique, including but not limited to the RNase protection method using riboprobes (Winter et al., Proc. Natl. Acad. Sci. USA 82:7575, 1985; Meyers et al., Science 230:1242, 1985) and proteins which recognize nucleotide mismatches, such as the E. coli mutS protein (Modrich, P. Ann. Rev. Genet. 25:229-253, 1991). Alternatively, variant alleles can be identified by single strand conformation polymorphism (SSCP) analysis (Orita et al., Genomics 5:874-879, 1989; Humphries et al., in Molecular Diagnosis of Genetic Diseases, R. Elles, ed., pp. 321-340, 1996) or denaturing gradient gel electrophoresis (DGGE) (Wartell et al., Nucl. Acids Res. 18:2699-2706, 1990; Sheffield et al., Proc. Natl. Acad. Sci. USA 86:232-236, 1989).

A polymerase-mediated primer extension method may also be used to identify the polymorphism(s). Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Patent 5,679,524. Related methods are disclosed in WO91/02087, WO90/09455, WO95/17676, U.S. Patent Nos. 5,302,509, and 5,945,283. Extended primers containing a polymorphism may be detected by mass spectrometry as described in U.S. Patent No. 5,605,798. Another primer extension method is allele-specific PCR (Ruaño et al., Nucl. Acids Res. 17:8392, 1989; Ruaño et al., Nucl. Acids Res. 19, 6877-6882, 1991; WO 93/22456; Turki et al., J. Clin. Invest. 95:1635-1641, 1995). In addition, multiple polymorphic sites may be investigated by simultaneously amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in Wallace et al. (WO89/10414).

In another aspect of the invention, an individual's TNFRSF11B haplotype pair is predicted from

its TNFRSF11B genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying a TNFRSF11B genotype for the individual at two or more polymorphic sites selected from PS1, PS2, PS3, PS4, PS5, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, and PS 25, enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing TNFRSF11B haplotype pairs identified in a reference population, and assigning a haplotype pair to the individual that is consistent with the data. In one embodiment, the reference haplotype pairs include the TNFRSF11B haplotype pairs shown in Table 4.

Generally, the reference population should be composed of randomly-selected individuals representing the major ethnogeographic groups of the world. A preferred reference population for use in the methods of the present invention comprises an approximately equal number of individuals from Caucasian, African American, Asian and Hispanic-Latino population groups with the minimum number of each group being chosen based on how rare a haplotype one wants to be guaranteed to see. For example, if one wants to have a q% chance of not missing a haplotype that exists in the population at a p% frequency of occurring in the reference population, the number of individuals (n) who must be sampled is given by $2n=\log(1-q)/\log(1-p)$ where p and q are expressed as fractions. A preferred reference population allows the detection of any haplotype whose frequency is at least 10% with about 99% certainty and comprises about 20 unrelated individuals from each of the four population groups named above. A particularly preferred reference population includes a 3-generation family representing one or more of the four population groups to serve as controls for checking quality of haplotyping procedures.

In a preferred embodiment, the haplotype frequency data for each ethnogeographic group is examined to determine whether it is consistent with Hardy-Weinberg equilibrium. Hardy-Weinberg equilibrium (D.L. Hartl et al., Principles of Population Genomics, Sinauer Associates (Sunderland, MA), 3rd Ed., 1997) postulates that the frequency of finding the haplotype pair H_1 / H_2 is equal to $p_{H-W}(H_1 / H_2) = 2p(H_1)p(H_2)$ if $H_1 \neq H_2$ and $p_{H-W}(H_1 / H_2) = p(H_1)p(H_2)$ if $H_1 = H_2$. A statistically significant difference between the observed and expected haplotype frequencies could be due to one or more factors including significant inbreeding in the population group, strong selective pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size does not reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., Nucleic Acids Res. 24:4841-4843, 1996).

In one embodiment of this method for predicting a TNFRSF11B haplotype pair, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the

reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., Nucleic Acids Res. 24:4841-4843, 1996).

The invention also provides a method for determining the frequency of a TNFRSF11B genotype or TNFRSF11B haplotype in a population. The method comprises determining the genotype or the haplotype pair for the TNFRSF11B gene that is present in each member of the population, wherein the genotype or haplotype comprises the nucleotide pair or nucleotide detected at one or more of the polymorphic sites PS1, PS2, PS3, PS4, PS5, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, and PS 25 in the TNFRSF11B gene; and calculating the frequency any particular genotype or haplotype is found in the population. The population may be a reference population, a family population, a same sex population, a population group, a trait population (e.g., a group of individuals exhibiting a trait of interest such as a medical condition or response to a therapeutic treatment).

In another aspect of the invention, frequency data for TNFRSF11B genotypes and/or haplotypes found in a reference population are used in a method for identifying an association between a trait and a TNFRSF11B genotype or a TNFRSF11B haplotype. The trait may be any detectable phenotype, including but not limited to susceptibility to a disease or response to a treatment. The method involves obtaining data on the frequency of the genotype(s) or haplotype(s) of interest in a reference population as well as in a population exhibiting the trait. Frequency data for one or both of the reference and trait populations may be obtained by genotyping or haplotyping each individual in the populations using one of the methods described above. The haplotypes for the trait population may be determined directly or, alternatively, by the predictive genotype to haplotype approach described above. In another embodiment, the frequency data for the reference and/or trait populations is obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is accessible by a computer. Once the frequency data is obtained, the frequencies of the genotype(s) or haplotype(s) of interest in the reference and trait populations are compared. In a preferred embodiment, the frequencies of all genotypes and/or haplotypes observed in the populations are compared. If a particular genotype or haplotype for the TNFRSF11B gene is more frequent in the trait population than in the reference population at a statistically significant amount, then the trait is predicted to be associated with that TNFRSF11B genotype or haplotype. Preferably, the

TNFRSF11B genotype or haplotype being compared in the trait and reference populations is selected from the full-genotypes and full-haplotypes shown in Tables 4 and 5, respectively, or from sub-genotypes and sub-haplotypes derived from these genotypes and haplotypes.

In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting TNFRSF11B or response to a therapeutic treatment for a medical condition. As used herein, "medical condition" includes but is not limited to any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment is desirable, and includes previously and newly identified diseases and other disorders. As used herein the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and adverse response (i.e., side effects).

In order to deduce a correlation between clinical response to a treatment and a TNFRSF11B genotype or haplotype, it is necessary to obtain data on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This clinical data may be obtained by analyzing the results of a clinical trial that has already been run and/or the clinical data may be obtained by designing and carrying out one or more new clinical trials. As used herein, the term "clinical trial" means any research study designed to collect clinical data on responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enroll subjects.

It is preferred that the individuals included in the clinical population have been graded for the existence of the medical condition of interest. This is important in cases where the symptom(s) being presented by the patients can be caused by more than one underlying condition, and where treatment of the underlying conditions are not the same. An example of this would be where patients experience breathing difficulties that are due to either asthma or respiratory infections. If both sets were treated with an asthma medication, there would be a spurious group of apparent non-responders that did not actually have asthma. These people would affect the ability to detect any correlation between haplotype and treatment outcome. This grading of potential patients could employ a standard physical exam or one or more lab tests. Alternatively, grading of patients could use haplotyping for situations where there is a strong correlation between haplotype pair and disease susceptibility or severity.

The therapeutic treatment of interest is administered to each individual in the trial population and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses and that the investigator will choose the number of responder groups (e.g., low, medium, high) made up by the various responses. In addition, the TNFRSF11B gene for each individual in the trial population is genotyped and/or haplotyped, which may be done before or after administering the treatment.

After both the clinical and polymorphism data have been obtained, correlations between individual response and TNFRSF11B genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their TNFRSF11B genotype or

haplotype (or haplotype pair) (also referred to as a polymorphism group), and then the averages and standard deviations of clinical responses exhibited by the members of each polymorphism group are calculated.

These results are then analyzed to determine if any observed variation in clinical response between polymorphism groups is statistically significant. Statistical analysis methods which may be used are described in L.D. Fisher and G. vanBelle, "Biostatistics: A Methodology for the Health Sciences", Wiley-Interscience (New York) 1993. This analysis may also include a regression calculation of which polymorphic sites in the TNFRSF11B gene give the most significant contribution to the differences in phenotype. One regression model useful in the invention starts with a model of the form

$$r = r_0 + S \times d$$

where r is the response, r_0 is a constant called the "intercept", S is the slope and d is the dose. To determine the dose, the most-common and least common nucleotides at the polymorphic site are first defined. Then, for each individual in the trial population, one calculates a "dose" as the number of least-common nucleotides the individual has at the polymorphic site of interest. This value can be 0 (homozygous for the least-common nucleotide), 1 (heterozygous), or 2 (homozygous for the most common nucleotide). An individual's "response" is the value of the clinical measurement. Standard linear regression methods are then used to fit all the individuals' doses and responses to a single model (see e.g., L.D. Fisher and G. vanBelle, supra, Ch 9). The outputs of the regression calculation are the intercept r_0, the slope S, and the variance (which measures how well the data fits this simple linear model). The Students t-test value and the level of significance can then be calculated for each of the polymorphic sites.

A second method for finding correlations between TNFRSF11B haplotype content and clinical responses uses predictive models based on error-minimizing optimization algorithms. One of many possible optimization algorithms is a genetic algorithm (R. Judson, "Genetic Algorithms and Their Uses in Chemistry" in Reviews in Computational Chemistry, Vol. 10, pp. 1-73, K. B. Lipkowitz and D. B. Boyd, eds. (VCH Publishers, New York, 1997). Simulated annealing (Press et al., "Numerical Recipes in C: The Art of Scientific Computing", Cambridge University Press (Cambridge) 1992, Ch. 10), neural networks (E. Rich and K. Knight, "Artificial Intelligence", 2nd Edition (McGraw-Hill, New York, 1991, Ch. 18), standard gradient descent methods (Press et al., supra Ch. 10), or other global or local optimization approaches (see discussion in Judson, supra) could also be used. As an example, a genetic algorithm approach is described herein. This method searches for optimal parameters or weights in linear or non-linear models connecting TNFRSF11B haplotype loci and clinical outcome. One model is of the form

$$C = C_0 + \sum_{\alpha}\left(\sum_{i} w_{i,\alpha} R_{i,\alpha} + \sum_{i} w'_{i,\alpha} L_{i,\alpha}\right) \qquad [1]$$

where C is the measured clinical outcome, i goes over all polymorphic sites, α over all candidate genes,

C_0, $w_{i,\alpha}$ and $w'_{i,\alpha}$ are variable weight values, $R_{i,\alpha}$ is equal to 1 if site i in gene α in the first haplotype takes on the most common nucleotide and –1 if it takes on the less common nucleotide. $L_{i,\alpha}$ is the same as $R_{i,\alpha}$ except for the second haplotype. The constant term C_0 and the weights $w_{i,\alpha}$ and $w'_{i,\alpha}$ are varied by the genetic algorithm during a search process that minimizes the error between the measured value of C and the value calculated from Equation 1. Models other than the one given in Equation 1 can be readily incorporated by those skilled in the art for analyzing the clinical and polymorphism data. The genetic algorithm is especially suited for searching not only over the space of weights in a particular model but also over the space of possible models (Judson, supra).

Correlations may also be analyzed using analysis of variation (ANOVA) techniques to determine how much of the variation in the clinical data is explained by different subsets of the polymorphic sites in the TNFRSF11B gene. ANOVA is used to test hypotheses about whether a response variable is caused by or correlated with one or more traits or variables that can be measured (Fisher and vanBelle, supra, Ch. 10). These traits or variables are called the independent variables. To carry out ANOVA, the independent variable(s) are measured and individuals are placed into groups based on their values for these variables. In this case, the independent variable(s) refers to the combination of polymorphisms present at a subset of the polymorphic sites, and thus, each group contains those individuals with a given genotype or haplotype pair. The variation in response within the groups and also the variation between groups is then measured. If the within-group response variation is large (people in a group have a wide range of responses) and the response variation between groups is small (the average responses for all groups are about the same) then it can be concluded that the independent variables used for the grouping are not causing or correlated with the response variable. For instance, if people are grouped by month of birth (which should have nothing to do with their response to a drug) the ANOVA calculation should show a low level of significance. However, if the response variation is larger between groups than within groups, the F-ratio (="between groups" divided by "within groups") is greater than one. Large values of the F-ratio indicate that the independent variable is causing or correlated with the response. The calculated F-ratio is preferably compared with the critical F-distribution value at whatever level of significance is of interest. If the F-ratio is greater than the Critical F-distribution value, then one may be confident that the individual's genotype or haplotype pair for this particular subset of polymorphic sites in the TNFRSF11B gene is at least partially responsible for, or is at least strongly correlated with the clinical response.

From the analyses described above, a mathematical model may be readily constructed by the skilled artisan that predicts clinical response as a function of TNFRSF11B genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the TNFRSF11B gene may be the basis for designing a diagnostic method to determine those individuals who will or will not respond to the treatment, or alternatively, will respond at

a lower level and thus may require more treatment, i.e., a greater dose of a drug. The diagnostic method may take one of several forms: for example, a direct DNA test (i.e., genotyping or haplotyping one or more of the polymorphic sites in the TNFRSF11B gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the diagnostic test results and the underlying TNFRSF11B genotype or haplotype that is in turn correlated with the clinical response. In a preferred embodiment, this diagnostic method uses the predictive haplotyping method described above.

Any or all analytical and mathematical operations involved in practicing the methods of the present invention may be implemented by a computer. In addition, the computer may execute a program that generates views (or screens) displayed on a display device and with which the user can interact to view and analyze large amounts of information relating to the TNFRSF11B gene and its genomic variation, including chromosome location, gene structure, and gene family, gene expression data, polymorphism data, genetic sequence data, and clinical data population data (e.g., data on ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations). The TNFRSF11B polymorphism data described herein may be stored as part of a relational database (e.g., an instance of an Oracle database or a set of ASCII flat files). These polymorphism data may be stored on the computer's hard drive or may, for example, be stored on a CD ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer via a network.

Preferred embodiments of the invention are described in the following examples. Other embodiments within the scope of the claims herein will be apparent to one skilled in the art from consideration of the specification or practice of the invention as disclosed herein. It is intended that the specification, together with the examples, be considered exemplary only, with the scope and spirit of the invention being indicated by the claims which follow the examples.

EXAMPLES

The Examples herein are meant to exemplify the various aspects of carrying out the invention and are not intended to limit the scope of the invention in any way. The Examples do not include detailed descriptions for conventional methods employed, such as in the performance of genomic DNA isolation, PCR and sequencing procedures. Such methods are well-known to those skilled in the art and are described in numerous publications, for example, Sambrook, Fritsch, and Maniatis, "Molecular Cloning: A Laboratory Manual", 2nd Edition, Cold Spring Harbor Laboratory Press, USA, (1989).

Example 1A

This example illustrates examination of various regions of the TNFRSF11B gene for polymorphic sites using DNA from Index Repository IA.

Amplification of Target Regions

The following target regions of the TNFRSF11B gene were amplified using the PCR primer pairs listed below, with the sequences presented in the 5′ to 3′ direction and nucleotide positions shown for each region corresponding to the indicated GenBank Accession No.

Accession Number: AB008821
Fragment 1
Forward Primer
599-618 GTACGGCGGAAACTCACAGC (SEQ ID NO:199)
Reverse Primer
Complement of 1190-1168 GCACAGCAACTTGTTCATTGTGG (SEQ ID NO:200)
PCR product 592 nt

Fragment 2
Forward Primer
864-883 TCTCCCAGGGGACAGACACC (SEQ ID NO:201)
Reverse Primer
Complement of 1389-1367 TAAAAGGACACCCTAGGGGAAGC (SEQ ID NO:202)
PCR product 526 nt

Fragment 3
Forward Primer
949-970 ATCCTGGAGCTTTCTGCACACC (SEQ ID NO:203)
Reverse Primer
Complement of 1609-1584 CTGAATCTAAGGGACCACTTCTTTGC (SEQ ID NO:204)
PCR product 661 nt

Accession Number: AB008822
Fragment 4
Forward Primer
95-118 GCTAAGATGATGCCACTGTGTTCC (SEQ ID NO:205)
Reverse Primer
Complement of 699-677 GCCCTGTAGTGGCAAAGTATTCC (SEQ ID NO:206)
PCR product 605 nt

Fragment 5
Forward Primer
4239-4262 CCTTTCCTCTCACATTTCATGAGC (SEQ ID NO:207)
Reverse Primer
Complement of 4976-4955 CAAACTTGACACTGCCCTTTGC (SEQ ID NO:208)
PCR product 738 nt

Fragment 6
Forward Primer
6526-6550 ACACGTAAGCTCAGTTGGTCTCTGC (SEQ ID NO:209)
Reverse Primer
Complement of 7144-7123 GAGCAGGGGTGAGAACAAAACC (SEQ ID NO:210)
PCR product 619 nt

Fragment 7
Forward Primer
8433-8455 CTTCCAAAGTTTTGTTGGATGCC (SEQ ID NO:211)
Reverse Primer
Complement of 9119-9096 AGTTTACTCATCCATGGGATCTCG (SEQ ID NO:212)

PCR product 687 nt

These primer pairs were used in PCR reactions containing genomic DNA isolated from immortalized cell lines for each member of Index Repository IA. The PCR reactions were carried out under the following conditions:

Reaction volume	= 20 µl
10 x Advantage 2 Polymerase reaction buffer (Clontech)	= 2 µl
100 ng of human genomic DNA	= 1 µl
10 mM dNTP	= 0.4 µl
Advantage 2 Polymerase enzyme mix (Clontech)	= 0.2 µl
Forward Primer (10 µM)	= 0.4 µl
Reverse Primer (10 µM)	= 0.4 µl
Water	=15.6µl

Amplification profile:
94°C - 2 min. 1 cycle

94°C - 30 sec.
70°C - 45 sec. 10 cycles
72°C - 1 min.

94°C - 30 sec.
64°C - 45 sec. 35 cycles
72°C - 1 min.

Sequencing of PCR Products

The PCR products were purified by Solid Phase Reversible Immobilization using the protocol developed by the Whitehead Genome Center. A detailed protocol can be found at http://www.genome.wi.mit.edu/sequencing/protocols/pure/SPRI_pcr.html.

Briefly, five µl of carboxyl coated magnetic beads (10 mg/ml) and 60 µl of HYB BUFFER (2.5M NaCl/20% PEG 8000) were added to each PCR reaction mixture (20 µl). The reaction mixture was mixed well and incubated at room temperature (RT) for 10 min. The microtitre plate was placed on a magnet for 2 min and the beads washed twice with 150 µl of 70% EtOH. The beads were air dried for 2 min and the DNA was eluted in 25 µl of distilled water and incubated at RT for 5 min. The beads were magnetically separated and the supernatant removed for testing and sequencing.

The purified PCR products were sequenced in both directions using the primer sets described previously or those listed, in the 5′ to 3′ direction, below.

Accession Number: AB008821
Fragment 1
Forward Primer
624-643 CCCAGCGAGAGGACAAAGGT (SEQ ID NO:213)
Reverse Primer
Complement of 1162-1144 GGAAACCTCAGGGGCTTGG (SEQ ID NO:214)

Fragment 2
Forward Primer

930-950 CCCAGCCCTGAAAGCGTTAAT (SEQ ID NO:215)
Reverse Primer
Complement of 1353-1334 AAAGCGGTTTCCTGCTCCAG (SEQ ID NO:216)

Fragment 3
Forward Primer
983-1002 CCGCCCAAGCTTCCTAAAAA (SEQ ID NO:217)
Reverse Primer
Complement of 1457-1436 TCTCCCTCTCTCTCGCTGTCTG (SEQ ID NO:218)

Accession Number: AB008822
Fragment 4
Forward Primer
151-170 AGTGGACCACCCAGGAAACG (SEQ ID NO:219)
Reverse Primer
Complement of 671-652 GCAATGGTCCTTTGGAAGCA (SEQ ID NO:220)

Fragment 5
Forward Primer
4314-4333 AACTGGCAAAGGGGATGATG (SEQ ID NO:221)
Reverse Primer
Complement of 4843-4822 GCATCGAGAGTAGCCTTAGCTG (SEQ ID NO:222)

Fragment 6
Forward Primer
6558-6577 ACCAGCCAACAGAAGCTTGA (SEQ ID NO:223)
Reverse Primer
Complement of 7069-7048 GTCCAACAATGATTCCAACAGG (SEQ ID NO:224)

Fragment 7
Forward Primer
8520-8541 GGTGTCACTTAACTCCCTCTCA (SEQ ID NO:225)
Reverse Primer
Complement of 9039-9020 CTGATTGGACCTGGTTACCT (SEQ ID NO:226)

Analysis of Sequences for Polymorphic Sites

Sequences were analyzed for the presence of polymorphisms using the Polyphred program (Nickerson et al., Nucleic Acids Res. 14:2745-2751, 1997). The presence of a polymorphism was confirmed on both strands. The polymorphisms and their locations in the TNFRSF11B gene are listed in Table 3 below.

Example 1B

This example illustrates examination of the OCIF gene for polymorphic sites in the following target regions: 2000 base pairs upstream of the ATG start codon; each of the exons, including approximately 100 base pairs on either side of the exon; and 500 to 1000 base pairs downstream of the termination codon.

Amplification of Target Regions

PCR primer pairs, which were designed based on the nearly complete OCIF genomic sequence

reported in the Genbank database (Accession Nos. AB008821 and AB008822), are set forth below:

Accession No. AB008821
Promoter fragment 1
Forward primer:
69-94 5'- CTGTAAACAATTTCAGTGGCAACCCG -3' (SEQ ID NO:227)
Reverse primer
427-402 5'- CCGTGCTATTCTGCATTCACTCCTTG -3' (SEQ ID NO:228)
PCR product 359 nt product

Promoter fragment 2
Forward primer:
263-288 5'- CGTAGGAAGCTCCGATACCAATAGCC -3' (SEQ ID NO:229)
Reverse primer:
868-846 5'- GGAGAGCAGGGGAAAAAAAAAGCC -3' (SEQ ID NO:230)
PCR product 606 nt

Promoter fragment 3
Forward primer:
558-583 5'- TTGAGGTTTCAGAACCCGAAGTGAAG -3' (SEQ ID NO:231)
Reverse primer
1069-1045 5'- CAACAGGAAGTATCGCCTGCCTTTG -3' (SEQ ID NO:232)
PCR product 512 nt

Exon 1
Forward primer:
945-970 5'- GTTAATCCTGGAGCTTTCTGCACACC -3' (SEQ ID NO:233)
Reverse primer:
1445-1420 5'- TCGCTGTCTGTCTCTCTCTTGCTGTC -3' (SEQ ID NO:234)
PCR product 501 nt

Accession No. AB008822
Exon 2
Forward primer:
91-114 5'- TCATGCTAAGATGATGCCACTGTG -3' (SEQ ID NO:235)
Reverse primer:
686-665 5'- GCAAAGTATTCCTCTGAGCAATGG -3' (SEQ ID NO:236)
PCR product: 598 nt

Exon 3
Forward primer:
4435-4457 5'- CTGTGTTAAGAGGGCATCTGCTG -3' (SEQ ID NO:237)
Reverse primer:
4791-4770 5'- TTGACCAAGAATGTGGCTGGAG -3' (SEQ ID NO:238)
PCR product: 357 nt

Exon 4
Forward primer:
6549-6573 5'- GCCACTAAGACCAGCCAACAGAAGC -3' (SEQ ID NO:239)
Reverse primer:
7166-7120 5'- GAGCAGGGGTGAGAACAAAACCTTG -3' (SEQ ID NO:240)
PCR product: 596 nt

Exon 5
Forward primer:

8150-8175 5'- GGCTGTGTGTCTCCTTTAGTTCCTCG -3' (SEQ ID NO:241)
Reverse primer:
9263-9238 5'- CTGACAGTTGGATTAACCATTTGGGG -3' (SEQ ID NO:242)
PCR product: 1114 nt

These primer pairs were used in PCR reactions containing genomic DNA isolated from immortalized cell lines for a reference population of 70 human individuals. The PCR reactions were carried out under the following conditions:

Reaction volume	=50 µl	
10 x Advantage 2 Polymerase reaction buffer (Clontech)		= 5 µl
100 ng of human genomic DNA 1	= 5 µl	
10 mM dNTP	= 1 µl	
Advantage 2 Polymerase enzyme mix (Clontech)	= 0.5 µl	
Forward Primer (10 µM)		= 1 µl
Reverse Primer (10 µM)		= 1 µl
Water		=36.5 µl

Amplification profile:

94°C - 2 min.	1 cycle
94°C - 30 sec. 70°C - 45 sec. 72°C - 1 min.	10 cycles
94°C - 30 sec. 64°C - 45 sec. 72°C - 1 min.	35 cycles

Sequencing of PCR Products

The PCR products were purified by Solid Phase Reversible Immobilization using the protocol developed by the Whitehead Genome Center. A detailed protocol can be found at http://www.genome.wi.mit.edu/sequencing/protocols/pure/SPRI_pcr.html.

Briefly, carboxyl coated magnetic beads (10 mg/ml) were washed three times with wash buffer (0.5 M EDTA, pH 8.0). Ten µl of washed beads and 50 µl of HYB BUFFER (2.5M NaCl/20% PEG 8000) were added to each PCR reaction mixture (50 µl). The reaction mixture was mixed well and incubated at RT for 10 min. The microtitre plate was placed on a magnet for 2 min and the beads washed twice with 150 µl of 70% EtOH. The beads were air dried for 2 min and resuspend in 20 µl of elution buffer (10 mM trisacetate, pH 7.8) and incubated at RT for 5 min. The beads were magnetically separated and the supernatant removed for testing and sequencing.

The purified PCR products were sequenced in both directions using sequencing primers that were identical to the PCR primers except for the exon 5 amplification product which, due to its size, was sequenced using two sets of primers. The Exon 5 sequencing primer sets are set forth below:

Accession No. AB008822
Forward primer 8150-8175
5'- GGCTGTGTGTCTCCTTTAGTTCCTCG -3' (SEQ ID NO:241)
Reverse primer 8928-8903

5'- TGGGAAAGTGGTACGTCTTTGAGTGC -3' (SEQ ID NO:243)

Forward primer 8674-8697
5'- TTGACCTCTGTGAAAACAGCGTGC -3' (SEQ ID NO:244)
Reverse primer 9145-9120
5'- CTGAAAGCCTCAAGTGCCTGAGAAAC -3' (SEQ ID NO:245)

Sequencing reactions were performed using the Big-Dye terminator kit from PE Biosystems (Foster City, CA) according to the manufacturer's instructions. The sequencing products were analyzed on an ABI 477 automated sequencer (PE Biosystems, Foster City, CA).

Analysis of Sequences for Polymorphic Sites

Sequences were analyzed for the presence of polymorphisms using the Polyphred program (Nickerson et al., 14 *Nucleic Acids Res.* 2745-2751, 1997). The presence of a polymorphism was confirmed on both the strands. The polymorphisms and their locations in the OCIF gene are listed in Table 3 below.

Table 3. Polymorphic sites Identified in the TNFRSF11B Gene

Polymorphic Site Number	Nucleotide Position	Reference Allele	Variant Allele	Example
PS1	491 (Acc. # AB008821)	G	A	1B
PS2	676 (Acc. # AB008821)	G	T	1A
PS3	889 (Acc. # AB008821)	C	T	1A
PS4	916 (Acc. # AB008821)	G	T	1A
PS5	950 (Acc. # AB008821)	T	C	1A
PS6[R]	1181 (Acc. # AB008821)	G	C	1A, 1B
PS7	1217 (Acc. # AB008821)	C	T	1A, 1B
PS8	1294 (Acc. # AB008821)	G	A	1A, 1B
PS9	1390 (Acc. # AB008821)	C	A	1A
PS10	505 (Acc. # AB008822)	C	T	1A
PS11	668 (Acc. # AB008822)	T	C	1A
PS12	4397 (Acc. # AB008822)	T	C	1A
PS13	4501 (Acc. # AB008822)	C	T	1A
PS14	6601 (Acc. # AB008822)	G	A	1A
PS15	6824 (Acc. # AB008822)	C	T	1A
PS16	6839 (Acc. # AB008822)	G	A	1A, 1B
PS17	6845 (Acc. # AB008822)	A	G	1A
PS18	6893 (Acc. # AB008822)	A	G	1A, 1B
PS19	6950 (Acc. # AB008822)	A	C	1B
PS20	8258 (Acc. # AB008822)	G	A	1B
PS21	8391 (Acc. # AB008822)	T	C	1B
PS22	8622 (Acc. # AB008822)	G	T	1B
PS23	8694 (Acc. # AB008822)	G	A	1A
PS24	8955 (Acc. # AB008822)	A	T	1B
PS25	9003 (Acc. # AB008822)	T	C	1A, 1B

[R]Previously reported in the literature.

Example 2

This example illustrates analysis of the TNFRSF11B polymorphisms identified in the Index

Repositories for human genotypes and haplotypes for all polymorphic sites except PS1 and PS16.

The different genotypes containing these polymorphisms that were observed in these reference populations are shown in Table 4 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below. In Table 4, homozygous positions are indicated by one nucleotide and heterozygous positions are indicated by two nucleotides.

TABLE 4A. GENOTYPES AND HAP PAIRS OBSERVED FOR THE TNFRSF11B GENE

Geno type	PS 2	PS 3	PS 4	PS 5	PS 6	PS 7	PS 8	PS 9	PS 10	PS 11	PS 12	PS 13	PS 14	PS 15	PS 17	PS 18	PS 19	PS 20	PS 21	PS 22	PS 23	PS 24	PS 25	Hap 1	Hap 2
1	G	C	G	C	C	C	G	C	C	T	C	C	G	C	A	A	A	G	T	G	G	A	T	1	1
2	G	C	G	T/C	C	C	G	C	C/T	T/C	T/C	C	G	C	A	A/G	A/C	A/G	T	G	G	A	T	1	2
3	G	C	G	T/C	G/C	C	G	C	C	T	T/C	C	G	C	A	A	A	G	T	G	G	A	T	1	5
4	G	C	G	T/C	G/C	C	G	C	C/T	T/C	T/C	C	G	C	A	G/A	A/C	A/G	T	G	G	A	T	1	6
5	G	C	G	C	G/C	C/T	G	C	C	T	T/C	C/T	G	C	A	A	A	G	T	G	G	A	T	1	9
6	G	C/T	G	C	C	C	G	C	C	T	C	C	G	C	A	A	A	A/G	T	G	G	A	T	1	12
7	G	C	G	T/C	G/C	C	G	C	C	T	T/C	C	G	C	A	A	A	G	T/C	G	G	A	T	1	13
8	G	C	G	C	C	C	G	C	C	T	C	C	G	C	A	A	A	G	T/C	G	G	A	T	1	14
9	G	C	G	T/C	G/C	C	G	C	C	T	T/C	C	G	C	A	A	A	G	T	G	G	A/T	T	1	15
10	G	C	G	T/C	C	C	G	C	C	T	T/C	C	G	C	A	A	A	G	T	G	G	A	T	1	18
11	G	C	G	C	C	C	G	C	C	T	C	C	G	C	A	A	A	G	T	G	G	A	T	1	23
12	G	C	G	C	C	C	G	C	C	T	T/C	C	G	C	A	A	A	G	T	G	G	A	T	1	25
13	G	C	G	C	G/C	C/T	G	C	C	T	T/C	C	G	C	A	A	A	G	T	G	G	A	T	1	27
14	G	C	G	T	C	C	G	C	T	C	T	C	G	C	A	G	C	A	T	G	G	A	T	2	2
15	G	C	G	T	G/C	C	G	C	C/T	T/C	T	C	G	C	A	A/G	A/C	A/G	T	G	G	A	T	2	5
16	G	C	G	T	G/C	C	G	C	T	C	T	C	G	C	A	G	C	A	T	G	G	A	T	2	6
17	G	C	G	T	G/C	C	G	C/A	C/T	T/C	T	C	G	C	A	A/G	A/C	A/G	T	G	G	A	T	2	19
18	G	C	G	T/C	G	C/T	A	C	C	T	T	C	G	C	A	A	A	G	T	G	G	A	C	3	7
19	G	C	G	T	G	C	G	C/A	C	T	T	C	G	C	A	A	A	G	T	G	G	A	T/C	4	5
20	G	C	G	T/C	G	C/T	G/A	C/A	C	T	T	C	G	C	A	A	A	G	T	G	G	A	C	4	7
21	G	C	G	T/C	G	C/T	G	C/A	C	T	T	C	G	C	A	A	A	G	T/C	G	G	A	C	4	8
22	G	C	G	T	G	C	G	C/A	C	T	T	C	G	C	A	A	A	G	T	G	G	A	C	4	21
23	G	C	G	T	G/C	C	G	C/A	C	T	T/C	C	G	C	A	A	A	G	T	G	G	A	T/C	4	23
24	G	C	G	T	G	C	G	C	C	T	T	C	G	C	A	A	A	G	T	G	G	A	T	5	5
25	G	C	G	T	G	C	G	C	C/T	T/C	T	C	G	C	A	G/A	A/C	A/G	T	G	G	A	T	5	6
26	G	C	G	T/C	G	C/T	G/A	C	C	T	T	C	G	C	A	A	A	G	T	G	G	A	T/C	5	7
27	G	C	G	T/C	G	C/T	G	C	C	T	T	C/T	G	C	A	A	A	G	T	G	G	A	T	5	9
28	G	C	G	T/C	G	C/T	G	C	C/T	T/C	T	C/T	G	C/T	A	A/G	A	G/A	T	G	G	A	T	5	10
29	G	C	G	T	G	C	G	C	C	T	T	C	G	C	A	A	A	G	T/C	G	G	A	T	5	13
30	G	C	G	T	G/C	C	G	C	C	T	T/C	C	G	C	A	A	A	G	T/C	G	G	A	T	5	14
31	G	C	G	T	G	C	G	C	C	T	T	C	G	C	A	A	A	G	T	G	G	A/T	T	5	15

TABLE 4B. GENOTYPES AND HAP PAIRS OBSERVED FOR THE TNFRSF11B GENE (CONTINUED)

Genotype	PS 2	PS 3	PS 4	PS 5	PS 6	PS 7	PS 8	PS 9	PS 10	PS 11	PS 12	PS 13	PS 14	PS 15	PS 17	PS 18	PS 19	PS 20	PS 21	PS 22	PS 23	PS 24	PS 25	Hap 1	Hap 2
32	G	C	G	T	G/C	C	G	C	C/T	T/C	T	C	G	C	A	G/A	A/C	A/G	T	G/T	G	A	T	5	16
33	G	C	G	T	G	C	G	C/A	C	T	T	C	G	C	A	A	A	G	T	G	G	A	T	5	19
34	G	C	G	T	G	C	G	C	C	T	T/C	C	G	C	A	A	A	G	T	G	G	A	T	5	20
35	G	C	G	T	G	C	G	C	C	T	T	C	G	C	A	A	A	G	T	G	G	A	T/C	5	21
36	G	C	G	T	G/C	C	G	C	C	T	T/C	C	G	C	A	A	A	G	T	G	G	A	T	5	23
37	G	C	G/T	T	G	C	G	C	C	T	T	C	G	C	A	A	A	G	T	G	G	A	T	5	24
38	G	C	G	T	G	C	G	C/A	C	T	T/C	C	G/A	C	A	A	A	G	T	G	G	A	T/C	5	26
39	G	C	G	T/C	G	C/T	G	C	C	T	T/C	C	G	C	A	A	A	G	T	G	G	A	T	5	27
40	G	C	G	T	G/C	C	G	C	C/T	T/C	T	C	G	C	A	A/G	A/C	A/G	T	G	G	A	T	6	18
41	G	C	G	T/C	G	C/T	A	C	C	T	T	C	G	C	A	A	A	G	T	G	G	A	T/C	7	11
42	G	C	G	T/C	G	C/T	G/A	C/A	C	T	T	C	G	C	A	A	A	G	T	G	G	A	T/C	7	19
43	G	C	G	T/C	G	C/T	G/A	C	C	T	T/C	C	G	C	A	A	A	G	T	G	G	A	T/C	7	20
44	G	C	G	T/C	G	C/T	G/A	C	C	T	T	C	G	C	A	A	A	G	T	G	G	A	C	7	21
45	G	C	G	C	G	T	G	C	C	T	T	T	G	C	A	A	A	G	T	G	G	A	T	9	9
46	G	C	G	T/C	G	C/T	G	C	C	T	T	C/T	G	C	A	A	A	G	T/C	G	G	A	T	9	13
47	G	C	G	T/C	G	C/T	G	C/A	C	T	T	C/T	G	C	A	A	A	G	T	G	G	A	T	9	19
48	G	C/T	G	C	C	C	G	C	C	T	C	C	G	C	A	A	A	A/G	T	G	G	A	T	12	23
49	G	C	G	T	G/C	C	G	C/A	C	T	T/C	C	G	C	A	A	A	G	T/C	G	G	A	T	14	19
50	G	C	G	T	G/C	C	G	C	C	T	T	C	G	C	A	A	A	G	T	G	G/A	A	T	17	23
51	G	C	G	T	G/C	C	G	C/A	C	T	T	C	G	C	A	A	A	G	T	G	G	A	T	18	19
52	G	C	G	T	G	C	G	A	C	T	T	C	G	C	A	A	A	G	T	G	G	A	T	19	19
53	G	C	G	T	G/C	C	G	C/A	C	T	T/C	C	G	C	A	A	A	G	T	G	G	A	T	19	23
54	G	C	G	T	G	C	G	C	C	T	T/C	C	G	C	A	A	A	G	T	G	G	A	T/C	20	21
55	G	C	G	T	G/C	C	G	C/A	C	T	T/C	C	G	C	A/G	A	A	G	T	G	G	A	T	22	23

The haplotype pairs shown in Table 4 were estimated from the unphased genotypes using an extension of Clark's algorithm (Clark, A.G. (1990) *Mol Bio Evol* 7, 111-122), as described in U.S. Provisional Patent Application filed April 19, 2000 and entitled "A Method and System for Determining Haplotypes from a Collection of Polymorphisms". In this method, haplotypes are assigned directly from individuals who are homozygous at all sites or heterozygous at no more than one of the variable sites. This list of haplotypes is then used to deconvolute the unphased genotypes in the remaining (multiply heterozygous) individuals.

By following this protocol, it was determined that the Index Repositories examined herein and, by extension, the general population contains the 27 human TNFRSF11B haplotypes shown in Table 5 below.

TABLE 5 HAPLOTYPES OBSERVED FOR TNFRSF11B GENE

HAP ID	PS 2	PS 3	PS 4	PS 5	PS 6	PS 7	PS 8	PS 9	PS 10	PS 11	PS 12	PS 13	PS 14	PS 15	PS 17	PS 18	PS 19	PS 20	PS 21	PS 22	PS 23	PS 24	PS 25
1	G	C	G	C	C	C	G	C	C	T	C	C	G	C	A	A	A	G	T	G	G	A	T
2	G	C	G	T	C	C	G	C	T	C	T	C	G	C	A	G	C	A	T	G	G	A	T
3	G	C	G	T	G	C	A	C	C	T	T	C	G	C	A	A	A	G	T	G	G	A	C
4	G	C	G	T	G	C	G	A	C	T	T	C	G	C	A	A	A	G	T	G	G	A	C
5	G	C	G	T	G	C	G	C	C	T	T	C	G	C	A	A	A	G	T	G	G	A	T
6	G	C	G	T	G	C	G	C	T	C	T	C	G	C	A	G	C	A	T	G	G	A	T
7	G	C	G	C	G	T	A	C	C	T	T	C	G	C	A	A	A	G	T	G	G	A	C
8	G	C	G	C	G	T	G	C	C	T	T	C	G	C	A	A	A	G	C	G	G	A	C
9	G	C	G	C	G	T	G	C	C	T	T	T	G	C	A	A	A	G	T	G	G	A	T
10	G	C	G	C	G	T	G	C	T	C	T	T	G	T	A	G	A	A	T	G	G	A	T
11	G	C	G	T	G	C	A	C	C	T	T	C	G	C	A	A	A	G	T	G	G	A	T
12	G	T	G	C	C	C	G	C	C	T	C	C	G	C	A	A	A	A	T	G	G	A	T
13	G	C	G	T	G	C	G	C	C	T	T	C	G	C	A	A	A	G	C	G	G	A	T
14	G	C	G	T	C	C	G	C	C	T	C	C	G	C	A	A	A	G	C	G	G	A	T
15	G	C	G	T	G	C	G	C	C	T	T	C	G	C	A	A	A	G	T	G	G	T	T
16	G	C	G	T	C	C	G	C	T	C	T	C	G	C	A	G	C	A	T	T	G	A	T
17	G	C	G	T	G	C	G	C	C	T	T	C	G	C	A	A	A	G	T	G	A	A	T
18	G	C	G	T	C	C	G	C	C	T	T	C	G	C	A	A	A	G	T	G	G	A	T
19	G	C	G	T	G	C	G	A	C	T	T	C	G	C	A	A	A	G	T	G	G	A	T
20	G	C	G	T	G	C	G	C	C	T	C	C	G	C	A	A	A	G	T	G	G	A	T
21	G	C	G	T	G	C	G	C	C	T	T	C	G	C	A	A	A	G	T	G	G	A	C
22	G	C	G	T	G	C	G	A	C	T	T	C	G	C	G	A	A	G	T	G	G	A	T
23	G	C	G	T	C	C	G	C	C	T	C	C	G	C	A	A	A	G	T	G	G	A	T
24	G	C	T	T	G	C	G	C	C	T	T	C	G	C	A	A	A	G	T	G	G	A	T
25	G	C	G	C	C	C	G	C	C	T	T	C	G	C	A	A	A	G	T	G	G	A	T
26	G	C	G	T	G	C	G	A	C	T	C	C	A	C	A	A	A	G	T	G	G	A	C
27	G	C	G	C	G	T	G	C	C	T	C	C	G	C	A	A	A	G	T	G	G	A	T

In view of the above, it will be seen that the several advantages of the invention are achieved and other advantageous results attained.

As various changes could be made in the above methods and compositions without departing from the scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

All references cited in this specification, including patents and patent applications, are hereby incorporated in their entirety by reference. The discussion of references herein is intended merely to summarize the assertions made by their authors and no admission is made that any reference constitutes prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

What is Claimed is:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:
   (a) a first nucleotide sequence which is a polymorphic variant of a reference sequence for Osteoclastogenesis Inhibitory Factor (TNFRSF11B) gene or a fragment thereof, wherein the reference sequence comprises SEQ ID NOS:1- 2, and the polymorphic variant comprises at least one polymorphism selected from the group consisting of adenine at PS1, thymine at PS2, thymine at PS3, thymine at PS4, cytosine at PS5, thymine at PS7, adenine at PS8, adenine at PS9, thymine at PS10, cytosine at PS11, cytosine at PS12, thymine at PS13, adenine at PS14, thymine at PS15, adenine at PS16, guanine at PS17, guanine at PS18, cytosine at PS19, adenine at PS20, cytosine at PS21, thymine at PS22, adenine at PS23, thymine at PS24 and cytosine at PS25; and
   (b) a second nucleotide sequence which is complementary to the first nucleotide sequence.
2. The isolated polynucleotide of claim 1 which comprises a TNFRSF11B isogene.
3. The isolated polynucleotide of claim 1 which is a DNA molecule and comprises both the first and second nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.
4. A recombinant organism transformed or transfected with the isolated polynucleotide of claim 1, wherein the organism expresses a TNFRSF11B protein encoded by the first nucleotide sequence.
5. The recombinant organism of claim 4 which is a nonhuman transgenic animal.
6. The isolated polynucleotide of claim 1, wherein the first nucleotide sequence is a polymorphic variant of a fragment of the TNFRSF11B gene, the fragment comprising one or more polymorphisms selected from the group consisting of adenine at PS1, thymine at PS2, thymine at PS3, thymine at PS4, cytosine at PS5, thymine at PS7, adenine at PS8, adenine at PS9, thymine at PS10, cytosine at PS11, cytosine at PS12, thymine at PS13, adenine at PS14, thymine at PS15, adenine at PS16, guanine at PS17, guanine at PS18, cytosine at PS19, adenine at PS20, cytosine at PS21, thymine at PS22, adenine at PS23, thymine at PS24 and cytosine at PS25.
7. An isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the TNFRSF11B cDNA or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:3 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of thymine at a position corresponding to nucleotide 699, adenine at a position corresponding to nucleotide 714, guanine at a position corresponding to nucleotide 720, guanine at a position corresponding to nucleotide 768, adenine at a position corresponding to nucleotide 841, thymine at a position corresponding to nucleotide 1102 and cytosine at a position corresponding to nucleotide 1150.
8. A recombinant organism transformed or transfected with the isolated polynucleotide of claim 7, wherein the organism expresses a Osteoclastogenesis Inhibitory Factor (TNFRSF11B) protein encoded by the polymorphic variant sequence.
9. The recombinant organism of claim 8 which is a nonhuman transgenic animal.

10. An isolated polypeptide comprising an amino acid sequence which is a polymorphic variant of a reference sequence for the TNFRSF11B protein or a fragment thereof, wherein the reference sequence comprises SEQ ID NO: 4 and the polymorphic variant comprises one or more variant amino acids selected from the group consisting of methionine at a position corresponding to amino acid position 240, methionine at a position corresponding to amino acid position 281, and serine at a position corresponding to amino acid 368.
11. An isolated antibody specific for and immunoreactive with the isolated polypeptide of claim 10.
12. A method for screening for drugs targeting the isolated polypeptide of claim 10 which comprises contacting the TNFRSF11B polymorphic variant with a candidate agent and assaying for binding activity.
13. A composition comprising at least one genotyping oligonucleotide for detecting a polymorphism in the Osteoclastogenesis Inhibitory Factor (TNFRSF11B) gene at a polymorphic site selected from PS1, PS2, PS3, PS4, PS5, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, and PS 25.
14. The composition of claim 13, wherein the genotyping oligonucleotide is an allele-specific oligonucleotide that specifically hybridizes to an allele of the TNFRSF11B gene at a region containing the polymorphic site.
15. The composition of claim 14, wherein the allele-specific oligonucleotide comprises a nucleotide sequence selected from the group consisting of of SEQ ID NOS:5-54, the complements of SEQ ID NOS: 5-54, and SEQ ID NOS:55 -150.
16. The composition of claim 13, wherein the genotyping oligonucleotide is a primer-extension oligonucleotide.
17. A method for genotyping the Osteoclastogenesis Inhibitory Factor (TNFRSF11B) gene of an individual, comprising determining for the two copies of the TNFRSF11B gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites (PS) selected from PS1, PS2, PS3, PS4, PS5, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, and PS 25.
18. The method of claim 17, wherein the determining step comprises:
    (a) isolating from the individual a nucleic acid mixture comprising both copies of the TNFRSF11B gene, or a fragment thereof, that are present in the individual;
    (b) amplifying from the nucleic acid mixture a target region containing at least one of the polymorphic sites;
    (c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;
    (d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

19. A method for haplotyping the Osteoclastogenesis Inhibitory Factor (TNFRSF11B) gene of an individual which comprises determining, for one copy of the TNFRSF11B gene present in the individual, the identity of the nucleotide at one or more polymorphic sites (PS) selected from PS1, PS2, PS3, PS4, PS5, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, and PS 25.

20. The method of claim 19, wherein the determining step comprises

(a) isolating from the individual a nucleic acid molecule containing only one of the two copies of the TNFRSF11B gene, or a fragment thereof, that is present in the individual;

(b) amplifying from the nucleic acid molecule a target region containing at least one of the polymorphic sites;

(c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;

(d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

21. A method for predicting a haplotype pair for the Osteoclastogenesis Inhibitory Factor (TNFRSF11B) gene of an individual comprising:

(a) identifying an TNFRSF11B genotype for the individual at two or more of polymorphic sites selected from PS1, PS2, PS3, PS4, PS5, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, and PS 25;

(b) enumerating all possible haplotype pairs which are consistent with the genotype;

(c) accessing data containing the TNFRSF11B haplotype pairs determined in a reference population; and

(d) assigning a haplotype pair to the individual that is consistent with the data.

22. A method for identifying an association between a trait and at least one genotype or haplotype of the Osteoclastogenesis Inhibitory Factor gene which comprises comparing the frequency of the genotype or haplotype in a population exhibiting the trait with the frequency of the genotype or haplotype in a reference population, wherein the genotype or haplotype comprises a nucleotide pair or nucleotide located at one or more polymorphic sites selected from PS1, PS2, PS3, PS4, PS5, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22, PS23, PS24, and PS 25, wherein a higher frequency of the genotype or haplotype in the trait population than in the reference population indicates the trait is associated with the genotype or haplotype.

23 The method of claim 22, wherein the haplotype is selected from haplotype numbers 1-27 shown in

Table 5.

24. The method of claim 23, wherein the trait is a clinical response to a drug targeting TNFRSF11B .

25. A computer system for storing and analyzing polymorphism data for the Osteoclastogenesis Inhibitory Factor gene, comprising:

(a) a central processing unit (CPU);

(b) a communication interface;

(c) a display device;

(d) an input device; and

(e) a database containing the polymorphism data;

wherein the polymorphism data comprises the genotypes and haplotype pairs shown in Table 4 and the haplotypes shown in Table 5.

26. A genome anthology for the Osteoclastogenesis Inhibitory Factor (TNFRSF11B) gene which comprises TNFRSF11B isogenes defined by haplotypes 1-27 shown in Table 5.

POLYMORPHISMS IN THE TNFRSF11B GENE (Accession No. AB008821)

```
CTGGAGACAT ATAACTTGAA CACTTGGCCC TGATGGGGAA GCAGCTCTGC
AGGGACTTTT TCAGCCATCT GTAAACAATT TCAGTGGCAA CCCGCGAACT       100
GTAATCCATG AATGGGACCA CACTTTACAA GTCATCAAGT CTAACTTCTA
GACCAGGGAA TTAATGGGGG AGACAGCGAA CCCTAGAGCA AAGTGCCAAA       200
CTTCTGTCGA TAGCTTGAGG CTAGTGGAAA GACCTCGAGG AGGCTACTCC
AGAAGTTCAG CGCGTAGGAA GCTCCGATAC CAATAGCCCT TTGATGATGG       300
TGGGGTTGGT GAAGGGAACA GTGCTCCGCA AGGTTATCCC TGCCCCAGGC
AGTCCAATTT TCACTCTGCA GATTCTCTCT GGCTCTAACT ACCCCAGATA       400
ACAAGGAGTG AATGCAGAAT AGCACGGGCT TTAGGGCCAA TCAGACATTA
GTTAGAAAAA TTCCTACTAC ATGGTTTATG TAAACTTGAA GATGAATGAT       500
                                              A
TGCGAACTCC CCGAAAAGGG CTCAGACAAT GCCATGCATA AAGAGGGGCC
CTGTAATTTG AGGTTTCAGA ACCCGAAGTG AAGGGGTCAG GCAGCCGGGT       600
ACGGCGGAAA CTCACAGCTT TCGCCCAGCG AGAGGACAAA GGTCTGGGAC
ACACTCCAAC TGCGTCCGGA TCTTGGCTGG ATCGGACTCT CAGGGTGGAG       700
                          T
GAGACACAAG CACAGCAGCT GCCCAGCGTG TGCCCAGCCC TCCCACCGCT
GGTCCCGGCT GCCAGGAGGC TGGCCGCTGG CGGGAAGGGG CCGGGAAACC       800
TCAGAGCCCC GCGGAGACAG CAGCCGCCTT GTTCCTCAGC CCGGTGGCTT
TTTTTTCCCC TGCTCTCCCA GGGGACAGAC ACCACCGCCC CACCCCTCAC       900
                                                  T
GCCCCACCTC CCTGGGGGAT CCTTTCCGCC CCAGCCCTGA AAGCGTTAAT
               T                                          C
CCTGGAGCTT TCTGCACACC CCCCGACCGC TCCCGCCCAA GCTTCCTAAA      1000
AAAGAAAGGT GCAAAGTTTG GTCCAGGATA GAAAAATGAC TGATCAAAGG
CAGGCGATAC TTCCTGTTGC CGGGACGCTA TATATAACGT GATGAGCGCA      1100
CGGGCTGCGG AGACGCACCG GAGCGCTCGC CCAGCCGCCG CCTCCAAGCC
CCTGAGGTTT CCGGGGACCA CAATGAACAA GTTGCTGTGC TGCGCGCTCG      1200
                                  C
TGGTAAGTCC CTGGGCCAGC CGACGGGTGC CCGGCGCCTG GGGAGGCTGC
               T
TGCCACCTGG TCTCCCAACC TCCCAGCGGA CCGGCGGGGA GAAGGCTCCA      1300
                                                 A
CTCGCTCCCT CCCAGGAGAG GCTTGGGGTT AGGCTGGAGC AGGAAACCGC
TTTCAAGTTA TGCCATGCTT CCCCTAGGGT GTCCTTTTAC GCTGCAAAGT      1400
                                          A
TCCTGCTGAC TTTATGGAAG ACAGCAAGAG AGAGACAGAC AGCGAGAGAG
AGGGAGAGAG AGAGAGAGAG AAACTTGTTT GAAAGTTTTA GTCATTAACC      1500
TTCTGTCTTC ATCTCAGAAT ATTAACGCCC TCATGTAGTC CATACTATCT
TTGCTTAATG AACTTGAACT TTTATTATTA GTGGCAAAGA AGTGGTCCCT      1600
TAGATTCAGA GTAAGTTGGA AGAAGACGTT AGTCTTCTTA AAACCATTAT
AATTAGAATA TGACATGATA GATTTTTCTA A                          1681
```

FIGURE 1A

POLYMORPHISMS IN THE TNFRSF11B GENE (Accession No. AB008822)

```
AAGCTTACTT TGTGCCAAAT CTCATTAGGC TTAAGGTAAT ACAGGACTTT
GAGTCAAATG ATACTGTTGC ACATAAGAAC AAACCTATTT TCATGCTAAG      100
ATGATGCCAC TGTGTTCCTT TCTCCTTCTA GTTTCTGGAC ATCTCCATTA
AGTGGACCAC CCAGGAAACG TTTCCTCCAA AGTACCTTCA TTATGACGAA      200
GAAACCTCTC ATCAGCTGTT GTGTGACAAA TGTCCTCCTG GTACCTACCT
AAAACAACAC TGTACAGCAA AGTGGAAGAC CGTGTGCGCC CCTTGCCCTG      300
ACCACTACTA CACAGACAGC TGGCACACCA GTGACGAGTG TCTATACTGC
AGCCCCGTGT GCAAGGAGCT GCAGTACGTC AAGCAGGAGT GCAATCGCAC      400
CCACAACCGC GTGTGCGAAT GCAAGGAAGG GCGCTACCTT GAGATAGAGT
TCTGCTTGAA ACATAGGAGC TGCCCTCCTG GATTTGGAGT GGTGCAAGCT      500
GGTACGTGTC AATGTGCAGC AAAATTAATT AGGATCATGC AAAGTCAGAT
    T
AGTTGTGACA GTTTAGGAGA ACACTTTTGT TCTGATGACA TTATAGGATA      600
GCAAATTGCA AAGGTAATGA AACCTGCCAG GTAGGTACTA TGTGTCTGGA
GTGCTTCCAA AGGACCATTG CTCAGAGGAA TACTTTGCCA CTACAGGGCA      700
                C
ATTTAATGAC AAATCTCAAA TGCAGCAAAT TATTCTCTCA TGAGATGCAT
GATGGTTTTT TTTTTTTTTT TTAAAGAAAC AAACTCAAGT TGCACTATTG      800
ATAGTTGATC TATACCTCTA TATTTCACTT CAGCATGGAC ACCTTCAAAC
TGCAGCACTT TTTGACAAAC ATCAGAAATG TTAATTTATA CCAAGAGAGT      900
AATTATGCTC ATATTAATGA GACTCTGGAG TGCTAACAAT AAGCAGTTAT
AATTAATTAT GTAAAAAAATG AGAATGGTGA GGGGAATTGC ATTTCATTAT     1000
TAAAAACAAG GCTAGTTCTT CCTTTAGCAT GGGAGCTGAG TGTTTGGGAG
GGTAAGGACT ATAGCAGAAT CTCTTCAATG AGCTTATTCT TTATCTTAGA     1100
CAAAACAGAT TGTCAAGCCA AGAGCAAGCA CTTGCCTATA AACCAAGTGC
TTTCTCTTTT GCATTTTGAA CAGCATTGGT CAGGGCTCAT GTGTATTGAA     1200
TCTTTTAAAC CAGTAACCCA CGTTTTTTTT CTGCCACATT TGCGAAGCTT
CAGTGCAGCC TATAACTTTT CATAGCTTGA GAAAATTAAG AGTATCCACT     1300
TACTTAGATG GAAGAAGTAA TCAGTATAGA TTCTGATGAC TCAGTTTGAA
GCAGTGTTTC TCAACTGAAG CCCTGCTGAT ATTTTAAGAA ATATCTGGAT     1400
TCCTAGGCTG GACTCCTTTT TGTGGGCAGC TGTCCTGCGC ATTGTAGAAT
TTTGGCAGCA CCCCTGGACT CTAGCCACTA GATACCAATA GCAGTCCTTC     1500
CCCCATGTGA CAGCCAAAAA TGTCTTCAGA CACTGTCAAA TGTCGCCAGG
TGGCAAAATC ACTCCTGGTT GAGAACAGGG TCATCAATGC TAAGTATCTG     1600
TAACTATTTT AACTCTCAAA ACTTGTGATA TACAAAGTCT AAATTATTAG
ACGACCAATA CTTTAGGTTT AAAGGCATAC AAATGAAACA TTCAAAAATC     1700
AAAATCTATT CTGTTTCTCA AATAGTGAAT CTTATAAAAT TAATCACAGA
AGATGCAAAT TGCATCAGAG TCCCTTAAAA TTCCTCTTCG TATGAGTATT     1800
TGAGGGAGGA ATTGGTGATA GTTCCTACTT TCTATTGGAT GGTACTTTGA
GACTCAAAAG CTAAGCTAAG TTGTGTGTGT GTCAGGGTGC GGGGTGTGGA     1900
ATCCCATCAG ATAAAAGCAA ATCCATGTAA TTCATTCAGT AAGTTGTATA
TGTAGAAAAA TGAAAAGTGG GCTATGCAGC TTGGAAACTA GAGAATTTTG     2000
AAAAATAATG GAAATCACAA GGATCTTTCT TAAATAAGTA AGAAAATCTG
TTTGTAGAAT GAAGCAAGCA GGCAGCCAGA AGACTCAGAA CAAAAGTACA     2100
CATTTTACTC TGTGTACACT GGCAGCACAG TGGGATTTAT TTACCTCTCC
CTCCCTAAAA ACCCACACAG CGGTTCCTCT TGGGAAATAA GAGGTTTCCA     2200
GCCCAAAGAG AAGGAAAGAC TATGTGGTGT TACTCTAAAA AGTATTTAAT
AACCGTTTTG TTGTTGCTGT TGCTGTTTTG AAATCAGATT GTCTCCTCTC     2300
CATATTTTAT TTACTTCATT CTGTTAATTC CTGTGGAATT ACTTAGAGCA
AGCATGGTGA ATTCTCAACT GTAAAGCCAA ATTTCTCCAT CATTATAATT     2400
TCACATTTTG CCTGGCAGGT TATAATTTTT ATATTTCCAC TGATAGTAAT
AAGGTAAAAT CATTACTTAG ATGGATAGAT CTTTTTCATA AAAAGTACCA     2500
TCAGTTATAG AGGGAAGTCA TGTTCATGTT CAGGAAGGTC ATTAGATAAA
GCTTCTGAAT ATATTATGAA ACATTAGTTC TGTCATTCTT AGATTCTTTT     2600
TGTTAAATAA CTTTAAAAGC TAACTTACCT AAAAGAAATA TCTGACACAT
```

FIGURE 2A

```
ATGAACTTCT CATTAGGATG CAGGAGAAGA CCCAAGCCAC AGATATGTAT      2700
CTGAAGAATG ÁACAAGATTC TTAGGCCCGG CACGGTGGCT CACATCTGTA
ATCTCAAGAG TTTGAGAGGT CAAGGCGGGC AGATCACCTG AGGTCAGGAG      2800
TTCAAGACCA GCCTGGCCAA CATGATGAAA CCCTGCCTCT ACTAAAAATA
CAAAAATTAG CAGGGCATGG TGGTGCATGC CTGCAACCCT AGCTACTCAG      2900
GAGGCTGAGA CAGGAGAATC TCTTGAACCC TCGAGGCGGA GGTTGTGGTG
AGCTGAGATC CCTCTACTGC ACTCCAGCCT GGGTGACAGA GATGAGACTC      3000
CGTCCCTGCC GCCGCCCCCG CCTTCCCCCC CAAAAAGATT CTTCTTCATG
CAGAACATAC GGCAGTCAAC AAAGGGAGAC CTGGGTCCAG GTGTCCAAGT      3100
CACTTATTTC GAGTAAATTA GCAATGAAAG AATGCCATGG AATCCCTGCC
CAAATACCTC TGCTTATGAT ATTGTAGAAT TTGATATAGA GTTGTATCCC      3200
ATTTAAGGAG TAGGATGTAG TAGGAAAGTA CTAAAAACAA ACACACAAAC
AGAAAACCCT CTTTGCTTTG TAAGGTGGTT CCTAAGATAA TGTCAGTGCA      3300
ATGCTGGAAA TAATATTTAA TATGTGAAGG TTTTAGGCTG TGTTTTCCCC
TCCTGTTCTT TTTTTCTGCC AGCCCTTTGT CATTTTTGCA GGTCAATGAA      3400
TCATGTAGAA AGAGACAGGA GATGAAACTA GAACCAGTCC ATTTTGCCCC
TTTTTTTATT TTCTGGTTTT GGTAAAAGAT ACAATGAGGT AGGAGGTTGA      3500
GATTTATAAA TGAAGTTTAA TAAGTTTCTG TAGCTTTGAT TTTTCTCTTT
CATATTTGTT ATCTTGCATA AGCCAGAATT GGCCTGTAAA ATCTACATAT      3600
GGATATTGAA GTCTAAATCT GTTCAACTAG CTTACACTAG ATGGAGATAT
TTTCATATTC AGATACACTG GAATGTATGA TCTAGCCATG CGTAATATAG      3700
TCAAGTGTTT GAAGGTATTT ATTTTTAATA GCGTCTTTAG TTGTGGACTG
GTTCAAGTTT TTCTGCCAAT GATTTCTTCA AATTTATCAA ATATTTTTCC      3800
ATCATGAAGT AAAATGCCCT TGCAGTCACC CTTCCTGAAG TTTGAACGAC
TCTGCTGTTT TAAACAGTTT AAGCAAATGG TATATCATCT TCCGTTTACT      3900
ATGTAGCTTA ACTGCAGGCT TACGCTTTTG AGTCAGCGGC CAACTTTATT
GCCACCTTCA AAAGTTTATT ATAATGTTGT AAATTTTTAC TTCTCAAGGT      4000
TAGCATACTT AGGAGTTGCT TCACAATTAG GATTCAGGAA AGAAAGAACT
TCAGTAGGAA CTGATTGGAA TTTAATGATG CAGCATTCAA TGGGTACTAA      4100
TTTCAAAGAA TGATATTACA GCAGACACAC AGCAGTTATC TTGATTTTCT
AGGAATAATT GTATGAAGAA TATGGCTGAC AACACGGCCT TACTGCCACT      4200
CAGCGGAGGC TGGACTAATG AACACCCTAC CCTTCTTTCC TTTCCTCTCA
CATTTCATGA GCGTTTTGTA GGTAACGAGA AAATTGACTT GCATTTGCAT      4300
TACAAGGAGG AGAAACTGGC AAAGGGGATG ATGGTGGAAG TTTTGTTCTG
TCTAATGAAG TGAAAAATGA AAATGCTAGA GTTTTGTGCA ACATAATAGT      4400
                                                    C
AGCAGTAAAA ACCAAGTGAA AAGTCTTTCC AAAACTGTGT TAAGAGGGCA
TCTGCTGGGA AACGATTTGA GGAGAAGGTA CTAAATTGCT TGGTATTTTC      4500
CGTAGGAACC ©CAGAGCGAA ATACAGTTTG CAAAAGATGT CCAGATGGGT
I
TCTTCTCAAA TGAGACGTCA TCTAAAGCAC CCTGTAGAAA ACACACAAAT      4600
TGCAGTGTGT TTGGTCTCCT GCTAACTCAG AAAGGAAATG CAACACACGA
CAACATATGT TCCGGAAACA GTGAATCAAC TCAAAAATGT GGAATAGGTA      4700
ATTACATTCC AAAATACGTC TTTGTACGAT TTTGTAGTAT CATCTCTCTC
TCTGAGTTGA ACACAAGGCC TCCAGCCACA TTCTTGGTCA AACTTACATT      4800
TTCCCTTTCT TGAATCTTAA CCAGCTAAGG CTACTCTCGA TGCATTACTG
CTAAAGCTAC CACTCAGAAT CTCTCAAAAA CTCATCTTCT CACAGATAAC      4900
ACCTCAAAGC TTGATTTTCT CTCCTTTCAC ACTGAAATCA AATCTTGCCC
ATAGGCAAAG GGCAGTGTCA AGTTTGCCAC TGAGATGAAA TTAGGAGAGT      5000
CCAAACTGTA GAATTCACGT TGTGTGTTAT TACTTTCACG AATGTCTGTA
TTATTAACTA AAGTATATAT TGGCAACTAA GAAGCAAAGT GATATAAACA      5100
TGATGACAAA TTAGGCCAGG CATGGTGGCT TACTCCTATA ATCCCAACAT
TTTGGGGGGC CAAGGTAGGC AGATCACTTG AGGTCAGGAT TTCAAGACCA      5200
GCCTGACCAA CATGGTGAAA CCTTGTCTCT ACTAAAAATA CAAAAATTAG
CTGGGCATGG TAGCAGGCAC TTCTAGTACC AGCTACTCAG GGCTGAGGCA      5300
GGAGAATCGC TTGAACCCAG GAGATGGAGG TTGCAGTGAG CTGAGATTGT
ACCACTGCAC TCCAGTCTGG GCAACAGAGC AAGATTTCAT CACACACACA      5400
```

FIGURE 2B

```
CACACACACA CACACACACA CATTAGAAAT GTGTACTTGG CTTTGTTACC
TATGGTATTA GTGCATCTAT TGCATGGAAC TTCCAAGCTA CTCTGGTTGT      5500
GTTAAGCTCT TCATTGGGTA CAGGTCACTA GTATTAAGTT CAGGTTATTC
GGATGCATTC CACGGTAGTG ATGACAATTC ATCAGGCTAG TGTGTGTGTT      5600
CACCTTGTCA CTCCCACCAC TAGACTAATC TCAGACCTTC ACTCAAAGAC
ACATTACACT AAAGATGATT TGCTTTTTTG TGTTTAATCA AGCAATGGTA      5700
TAAACCAGCT TGACTCTCCC CAAACAGTTT TTCGTACTAC AAAGAAGTTT
ATGAAGCAGA GAAATGTGAA TTGATATATA TATGAGATTC TAACCCAGTT      5800
CCAGCATTGT TTCATTGTGT AATTGAAATC ATAGACAAGC CATTTTAGCC
TTTGCTTTCT TATCTAAAAA AAAAAAAAAA AAAATGAAGG AAGGGGTATT      5900
AAAAGGAGTG ATCAAATTTT AACATTCTCT TTAATTAATT CATTTTTAAT
TTTACTTTTT TTCATTTATT GTGCACTTAC TATGTGGTAC TGTGCTATAG      6000
AGGCTTTAAC ATTTATAAAA ACACTGTGAA AGTTGCTTCA GATGAATATA
GGTAGTAGAA CGGCAGAACT AGTATTCAAA GCCAGGTCTG ATGAATCCAA      6100
AAACAAACAC CCATTACTCC CATTTTCTGG GACATACTTA CTCTACCCAG
ATGCTCTGGG CTTTGTAATG CCTATGTAAA TAACATAGTT TTATGTTTGG      6200
TTATTTTCCT ATGTAATGTC TACTTATATA TCTGTATCTA TCTCTTGCTT
TGTTTCCAAA GGTAAACTAT GTGTCTAAAT GTGGGCAAAA AATAACACAC      6300
TATTCCAAAT TACTGTTCAA ATTCCTTTAA GTCAGTGATA ATTATTTGTT
TTGACATTAA TCATGAAGTT CCCTGTGGGT ACTAGGTAAA CCTTTAATAG      6400
AATGTTAATG TTTGTATTCA TTATAAGAAT TTTTGGCTGT TACTTATTTA
CAACAATATT TCACTCTAAT TAGACATTTA CTAAACTTTC TCTTGAAAAC      6500
AATGCCCAAA AAAGAACATT AGAAGACACG TAAGCTCAGT TGGTCTCTGC
CACTAAGACC AGCCAACAGA AGCTTGATTT TATTCAAACT TTGCATTTTA      6600
GCATATTTTA TCTTGGAAAA TTCAATTGTG TTGGTTTTTT GTTTTTGTTT
A
GTATTGAATA GACTCTCAGA AATCCAATTG TTGAGTAAAT CTTCTGGGTT      6700
TTCTAACCTT TCTTTAGATG TTAGCCTGTG TGAGGAGGCA TTCTTCAGGT
TTGCTGTTCC TACAAAGTTT ACGCCTAACT GGCTTAGTGT CTTGGTAGAC      6800
AATTTGCCTG GCACCAAAGT AAACGCAGAG AGTGTAGAGA GGATAAAACG
                    T           A       G
GCAACACAGC TCACAAGAAC AGACTTTCCA GCTGCTGAAG TTATGGAAAC      6900
                                          G
ATCAAAACAA AGACCAAGAT ATAGTCAAGA AGATCATCCA AGGTATGATA
                                            C
ATCTAAAATA AAAAGATCAA TCAGAAATCA AAGACACCTA TTTATCATAA      7000
ACCAGGAACA AGACTGCATG TATGTTTAGT TGTGTGGATC TTGTTTCCCT
GTTGGAATCA TTGTTGGACT GAAAAAGTTT CCACCTGATA ATGTAGATGT      7100
GATTCCACAA ACAGTTATAC AAGGTTTTGT TCTCACCCCT GCTCCCCAGT
TTCCTTGTAA AGTATGTTGA ACACTCTAAG AGAAGAGAAA TGCATTTGAA      7200
GGCAGGGCTG TATCTCAGGG AGTCGCTTCC AGATCCCTTA ACGCTTCTGT
AAGCAGCCCC TCTAGACCAC CAAGGAGAAG CTCTATAACC ACTTTGTATC      7300
TTACATTGCA CCTCTACCAA GAAGCTCTGT TGTATTTACT TGGTAATTCT
CTCCAGGTAG GCTTTTCGTA GCTTACAAAT ATGTTCTTAT TAATCCTCAT      7400
GATATGGCCT GCATTAAAAT TATTTTAATG GCATATGTTA TGAGAATTAA
TGAGATAAAA TCTGAAAAGT GTTTGAGCCT CTTGTAGGAA AAAGCTAGTT      7500
ACAGCAAAAT GTTCTCACAT CTTATAAGTT TATATAAAGA TTCTCCTTTA
GAAATGGTGT GAGAGAGAAA CAGAGAGAGA TAGGGAGAGA AGTGTGAAAG      7600
AATCTGAAGA AAAGGAGTTT CATCCAGTGT GGACTGTAAG CTTTACGACA
CATGATGGAA AGAGTTCTGA CTTCAGTAAG CATTGGGAGG ACATGCTAGA      7700
AGAAAAAGGA AGAAGAGTTT CCATAATGCA GACAGGGTCA GTGAGAAATT
CATTCAGGTC CTCACCAGTA GTTAAATGAC TGTATAGTCT TGCACTACCC      7800
TAAAAAACTT CAAGTATCTG AAACCGGGGC AACAGATTTT AGGAGACCAA
CGTCTTTGAG AGCTGATTGC TTTTGCTTAT GCAAAGAGTA AACTTTTATG      7900
TTTTGAGCAA ACCAAAAGTA TTCTTTGAAC GTATAATTAG CCCTGAAGCC
GAAAGAAAAG AGAAAATCAG AGACCGTTAG AATTGGAAGC AACCAAATTC      8000
CCTATTTTAT AAATGAGGAC ATTTTAACCC AGAAAGATGA ACCGATTTGG
```

FIGURE 2C

```
CTTAGGGCTC ACAGATACTA AGTGACTCAT GTCATTAATA GAAATGTTAG      8100
TTCCTCCCTC TTAGGTTTGT ACCCTAGCTT ATTACTGAAA TATTCTCTAG
GCTGTGTGTC TCCTTTAGTT CCTCGACCTC ATGTCTTTGA GTTTTCAGAT      8200
ATCCTCCTCA TGGAGGTAGT CCTCTGGTGC TATGTGTATT CTTTAAAGGC
TAGTTACGGC AATTAACTTA TCAACTAGCG CCTACTAATG AAACTTTGTA      8300
       A
TTACAAAGTA GCTAACTTGA ATACTTTCCT TTTTTTCTGA AATGTTATGG
TGGTAATTTC TCAAACTTTT TCTTAGAAAA CTGAGAGTGA TGTGTCTTAT      8400
                                            C
TTTCTACTGT TAATTTTCAA AATTAGGAGC TTCTTCCAAA GTTTTGTTGG
ATGCCAAAAA TATATAGCAT ATTATCTTAT TATAACAAAA AATATTTATC      8500
TCAGTTCTTA GAAATAAATG GTGTCACTTA ACTCCCTCTC AAAAGAAAAG
GTTATCATTG AAATATAATT ATGAAATTCT GCAAGAACCT TTTGCCTCAC      8600
GCTTGTTTTA TGATGGCATT GGATGAATAT AAATGATGTG AACACTTATC
                       T
TGGGCTTTTG CTTTATGCAG ATATTGACCT CTGTGAAAAC AGCGTGCAGC      8700
                                              A
GGCACATTGG ACATGCTAAC CTCACCTTCG AGCAGCTTCG TAGCTTGATG
GAAAGCTTAC CGGGAAAGAA AGTGGGAGCA GAAGACATTG AAAAAACAAT      8800
AAAGGCATGC AAACCCAGTG ACCAGATCCT GAAGCTGCTC AGTTTGTGGC
GAATAAAAAA TGGCGACCAA GACACCTTGA AGGGCCTAAT GCACGCACTA      8900
AAGCACTCAA AGACGTACCA CTTTCCCAAA ACTGTCAGTC AGAGTCTAAA
GAAGACCATC AGGTTCCTTC ACAGCTTCAC AATGTACAAA TTGTATCAGA      9000
  T
AGTTATTTTT AGAAATGATA GGTAACCAGG TCCAATCAGT AAAAATAAGC
  C
TGCTTATAAC TGGAAATGGC CATTGAGCTG TTTCCTCACA ATTGGCGAGA      9100
TCCCATGGAT GAGTAAACTG TTTCTCAGGC ACTTGAGGCT TTCAGTGATA
TCTTTCTCAT TACCAGTGAC TAATTTTGCC ACAGGGTACT AAAAGAAACT      9200
ATGATGTGGA GAAAGGACTA ACATCTCCTC CAATAAACCC CAAATGGTTA
ATCCAACTGT CAGATCTGGA TCGTTATCTA CTGACTATAT TTTCCCTTAT      9300
TACTGCTTGC AGTAATTCAA CTGGAAATTA AAAAAAAAAA ACTAGACTCC
ACTGGGCCTT ACTAAATATG GGAATGTCTA ACTTAAATAG CTTTGGGATT      9400
CCAGCTATGC TAGAGGCTTT TATTAGAAAG CCATATTTTT TTCTGTAAAA
GTTACTAATA TATCTGTAAC ACTATTACAG TATTGCTATT TATATTCATT      9500
CAGATATAAG ATTTGGACAT ATTATCATCC TATAAAGAAA CGGTATGACT
TAATTTTAGA AAGAAAATTA TATTCTGTTT ATTATGACAA ATGAAAGAGA      9600
AAATATATAT TTTTAATGGA AAGTTTGTAG CATTTTTCTA ATAGGTACTG
CCATATTTTT CTGTGTGGAG TATTTTTATA ATTTTATCTG TATAAGCTGT      9700
AATATCATTT TATAGAAAAT GCATTATTTA GTCAATTGTT TAATGTTGGA
AAACATATGA AATATAAATT ATCTGAATAT TAGATGCTCT GAGAAATTGA      9800
ATGTACCTTA TTTAAAAGAT TTTATGGTTT TATAACTATA TAAATGACAT
TATTAAAGTT TTCAAATTA                                        9869
```

FIGURE 2D

POLYMORPHISMS IN THE CODING SEQUENCE OF TNFRSF11B

```
ATGAACAAGT TGCTGTGCTG CGCGCTCGTG TTTCTGGACA TCTCCATTAA
       C
GTGGACCACC CAGGAAACGT TTCCTCCAAA GTACCTTCAT TATGACGAAG      100
AAACCTCTCA TCAGCTGTTG TGTGACAAAT GTCCTCCTGG TACCTACCTA
AAACAACACT GTACAGCAAA GTGGAAGACC GTGTGCGCCC CTTGCCCTGA      200
CCACTACTAC ACAGACAGCT GGCACACCAG TGACGAGTGT CTATACTGCA
GCCCCGTGTG CAAGGAGCTG CAGTACGTCA AGCAGGAGTG CAATCGCACC      300
CACAACCGCG TGTGCGAATG CAAGGAAGGG CGCTACCTTG AGATAGAGTT
CTGCTTGAAA CATAGGAGCT GCCCTCCTGG ATTTGGAGTG GTGCAAGCTG      400
GAACCCCAGA GCGAAATACA GTTTGCAAAA GATGTCCAGA TGGGTTCTTC
TCAAATGAGA CGTCATCTAA AGCACCCTGT AGAAAACACA CAAATTGCAG      500
TGTCTTTGGT CTCCTGCTAA CTCAGAAAGG AAATGCAACA CACGACAACA
TATGTTCCGG AAACAGTGAA TCAACTCAAA AATGTGGAAT AGATGTTACC      600
CTGTGTGAGG AGGCATTCTT CAGGTTTGCT GTTCCTACAA AGTTTACGCC
TAACTGGCTT AGTGTCTTGG TAGACAATTT GCCTGGCACC AAAGTAAACG      700
                                                    T
CAGAGAGTGT AGAGAGGATA AAACGGCAAC ACAGCTCACA AGAACAGACT
           . A   G
TTCCAGCTGC TGAAGTTATG GAAACATCAA AACAAAGACC AAGATATAGT      800
             G
CAAGAAGATC ATCCAAGATA TTGACCTCTG TGAAAACAGC GTGCAGCGGC
                                           A
ACATTGGACA TGCTAACCTC ACCTTCGAGC AGCTTCGTAG CTTGATGGAA      900
AGCTTACCGG GAAAGAAAGT GGGAGCAGAA GACATTGAAA AAACAATAAA
GGCATGCAAA CCCAGTGACC AGATCCTGAA GCTGCTCAGT TTGTGGCGAA     1000
TAAAAAATGG CGACCAAGAC ACCTTGAAGG GCCTAATGCA CGCACTAAAG
CACTCAAAGA CGTACCACTT TCCCAAAACT GTCACTCAGA GTCTAAAGAA     1100
GACCATCAGG TTCCTTCACA GCTTCACAAT GTACAAATTG TATCAGAAGT
 T                                                   C
TATTTTTAGA AATGATAGGT AACCAGGTCC AATCAGTAAA AATAAGCTGC     1200
TTATAA                                                    1206
```

FIGURE 3

ISOFORMS OF THE TNFRSF11B PROTEIN

```
MNKLLCCALV FLDISIKWTT QETFPPKYLH YDEETSHQLL CDKCPPGTYL
  N
KQHCTAKWKT VCAPCPDHYY TDSWHTSDEC LYCSPVCKEL QYVKQECNRT      100
HNRVCECKEG RYLEIEFCLK HRSCPPGFGV VQAGTPERNT VCKRCPDGFF
SNETSSKAPC RKHTNCSVFG LLLTQKGNAT HDNICSGNSE STQKCGIDVT      200
LCEEAFFRFA VPTKFTPNWL SVLVDNLPGT KVNAESVERI KRQHSSQEQT
                                          M
FQLLKLWKHQ NKDQDIVKKI IQDIDLCENS VQRHIGHANL TFEQLRSLME      300
                                  M
SLPGKKVGAE DIEKTIKACK PSDQILKLLS LWRIKNGDQD TLKGLMHALK
HSKTYHFPKT VTQSLKKTIR FLHSFTMYKL YQKLFLEMIG NQVQSVKISC      400
                  S
L                                                            401
```

FIGURE 4

SEQUENCE LISTING

<110> Genaissance Pharmaceuticals, Inc.
Chew, Anne
Denton, R. Rex
Duda, Amy
Nandabalan, Krishnan
Stevens, J. Claiborne

<120> Drug Target Isogenes: Polymorphisms in the Osteoclastogenesis Inhibitory Factor

<130> MWH-001PCT TNFRSF11B-OCIF

<140> TBA
<141> 2000-07-10

<150> 60/143,020
<151> 1999-07-09

<160> 245

<170> PatentIn Ver. 2.1

<210> 1
<211> 1681
<212> DNA
<213> Homo sapiens

<400> 1
ctggagacat ataacttgaa cacttggccc tgatggggaa gcagctctgc agggactttt 60
tcagccatct gtaaacaatt tcagtggcaa cccgcgaact gtaatccatg aatgggacca 120
cactttacaa gtcatcaagt ctaacttcta gaccagggaa ttaatggggg agacagcgaa 180
ccctagagca aagtgccaaa cttctgtcga tagcttgagg ctagtggaaa gacctcgagg 240
aggctactcc agaagttcag cgcgtaggaa gctccgatac caatagccct ttgatgatgg 300
tggggttggt gaagggaaca gtgctccgca aggttatccc tgccccaggc agtccaattt 360
tcactctgca gattctctct ggctctaact accccagata acaaggagtg aatgcagaat 420
agcacgggct ttagggccaa tcagacatta gttagaaaaa ttcctactac atggtttatg 480
taaacttgaa gatgaatgat tgcgaactcc ccgaaaaggg ctcagacaat gccatgcata 540
aagaggggcc ctgtaatttg aggtttcaga acccgaagtg aaggggtcag gcagccgggt 600
acggcggaaa ctcacagctt tcgcccagcg agaggacaaa ggtctgggac acactccaac 660
tgcgtccgga tcttggctgg atcggactct cagggtggag gagacacaag cacagcagct 720
gcccagcgtg tgcccagccc tcccaccgct ggtcccggct gccaggaggc tggccgctgg 780
cgggaagggg ccgggaaacc tcagagcccc gcggagacag cagccgcctt gttcctcagc 840
ccggtggctt ttttttcccc tgctctccca ggggacagac accaccgccc caccctcac 900
gccccacctc cctgggggat cctttccgcc ccagccctga aagcgttaat cctggagctt 960
tctgcacacc ccccgaccgc tcccgcccaa gcttcctaaa aaagaaaggt gcaaagtttg 1020
gtccaggata gaaaaatgac tgatcaaagg caggcgatac ttcctgttgc cgggacgcta 1080

tatataacgt gatgagcgca cgggctgcgg agacgcaccg gagcgctcgc ccagccgccg 1140
cctccaagcc cctgaggttt ccggggacca caatgaacaa gttgctgtgc tgcgcgctcg 1200
tggtaagtcc ctgggccagc cgacgggtgc ccggcgcctg gggaggctgc tgccacctgg 1260
tctcccaacc tcccagcgga ccggcgggga gaaggctcca ctcgctccct cccaggagag 1320
gcttggggtt aggctggagc aggaaaccgc tttcaagtta tgccatgctt cccctagggt 1380
gtccttttac gctgcaaagt tcctgctgac tttatggaag acagcaagag agagacagac 1440
agcgagagag agggagagag agagagagag aaacttgttt gaaagtttta gtcattaacc 1500
ttctgtcttc atctcagaat attaacgccc tcatgtagtc catactatct ttgcttaatg 1560
aacttgaact tttattatta gtggcaaaga agtggtccct tagattcaga gtaagttgga 1620
agaagacgtt agtcttctta aaaccattat aattagaata tgacatgata gatttttcta 1680
a 1681

<210> 2
<211> 9869
<212> DNA
<213> Homo sapiens

<400> 2
aagcttactt tgtgccaaat ctcattaggc ttaaggtaat acaggacttt gagtcaaatg 60
atactgttgc acataagaac aaacctattt tcatgctaag atgatgccac tgtgttcctt 120
tctccttcta gtttctggac atctccatta agtggaccac ccaggaaacg tttcctccaa 180
agtaccttca ttatgacgaa gaaacctctc atcagctgtt gtgtgacaaa tgtcctcctg 240
gtacctacct aaaacaacac tgtacagcaa agtggaagac cgtgtgcgcc ccttgccctg 300
accactacta cacagacagc tggcacacca gtgacgagtg tctatactgc agccccgtgt 360
gcaaggagct gcagtacgtc aagcaggagt gcaatcgcac ccacaaccgc gtgtgcgaat 420
gcaaggaagg gcgctacctt gagatagagt tctgcttgaa acataggagc tgccctcctg 480
gatttggagt ggtgcaagct ggtacgtgtc aatgtgcagc aaaattaatt aggatcatgc 540
aaagtcagat agttgtgaca gtttaggaga acacttttgt tctgatgaca ttataggata 600
gcaaattgca aaggtaatga aacctgccag gtaggtacta tgtgtctgga gtgcttccaa 660
aggaccattg ctcagaggaa tactttgcca ctacagggca atttaatgac aaatctcaaa 720
tgcagcaaat tattctctca tgagatgcat gatggttttt ttttttttt ttaaagaaac 780
aaactcaagt tgcactattg atagttgatc tatacctcta tatttcactt cagcatggac 840
accttcaaac tgcagcactt tttgacaaac atcagaaatg ttaatttata ccaagagagt 900
aattatgctc atattaatga gactctggag tgctaacaat aagcagttat aattaattat 960
gtaaaaaatg agaatggtga ggggaattgc atttcattat taaaaacaag gctagttctt 1020
cctttagcat gggagctgag tgtttgggag ggtaaggact atagcagaat ctcttcaatg 1080
agcttattct ttatcttaga caaaacagat tgtcaagcca agagcaagca cttgcctata 1140
aaccaagtgc tttctctttt gcattttgaa cagcattggt cagggctcat gtgtattgaa 1200
tcttttaaac cagtaaccca cgttttttt ctgccacatt tgcgaagctt cagtgcagcc 1260
tataactttt catagcttga gaaaattaag agtatccact tacttagatg gaagaagtaa 1320
tcagtataga ttctgatgac tcagtttgaa gcagtgtttc tcaactgaag ccctgctgat 1380
attttaagaa atatctggat tcctaggctg gactcctttt tgtgggcagc tgtcctgcgc 1440
attgtagaat tttggcagca cccctggact ctagccacta gataccaata gcagtccttc 1500
ccccatgtga cagccaaaaa tgtcttcaga cactgtcaaa tgtcgccagg tggcaaaatc 1560
actcctggtt gagaacaggg tcatcaatgc taagtatctg taactatttt aactctcaaa 1620
acttgtgata tacaaagtct aaattattag acgaccaata ctttaggttt aaaggcatac 1680
aaatgaaaca ttcaaaaatc aaaatctatt ctgtttctca aatagtgaat cttataaaat 1740

taatcacaga agatgcaaat tgcatcagag tcccttaaaa ttcctcttcg tatgagtatt 1800
tgagggagga attggtgata gttcctactt tctattggat ggtactttga gactcaaaag 1860
ctaagctaag ttgtgtgtgt gtcagggtgc ggggtgtgga atcccatcag ataaaagcaa 1920
atccatgtaa ttcattcagt aagttgtata tgtagaaaaa tgaaaagtgg gctatgcagc 1980
ttggaaacta gagaattttg aaaaataatg gaaatcacaa ggatctttct taaataagta 2040
agaaaatctg tttgtagaat gaagcaagca ggcagccaga agactcagaa caaaagtaca 2100
cattttactc tgtgtacact ggcagcacag tgggatttat ttacctctcc ctcccctaaaa 2160
acccacacag cggttcctct tgggaaataa gaggtttcca gcccaaagag aaggaaagac 2220
tatgtggtgt tactctaaaa agtatttaat aaccgttttg ttgttgctgt tgctgttttg 2280
aaatcagatt gtctcctctc catattttat ttacttcatt ctgttaattc ctgtggaatt 2340
acttagagca agcatggtga attctcaact gtaaagccaa atttctccat cattataatt 2400
tcacattttg cctggcaggt tataattttt atatttccac tgatagtaat aaggtaaaat 2460
cattacttag atggatagat ctttttcata aaaagtacca tcagttatag agggaagtca 2520
tgttcatgtt caggaaggtc attagataaa gcttctgaat atattatgaa acattagttc 2580
tgtcattctt agattctttt tgttaaataa ctttaaaagc taacttacct aaaagaaata 2640
tctgacacat atgaacttct cattaggatg caggagaaga cccaagccac agatatgtat 2700
ctgaagaatg aacaagattc ttaggcccgg cacggtggct cacatctgta atctcaagag 2760
tttgagaggt caaggcgggc agatcacctg aggtcaggag ttcaagacca gcctggccaa 2820
catgatgaaa ccctgcctct actaaaaata caaaaattag cagggcatgg tggtgcatgc 2880
ctgcaaccct agctactcag gaggctgaga caggagaatc tcttgaaccc tcgaggcgga 2940
ggttgtggtg agctgagatc cctctactgc actccagcct gggtgacaga gatgagactc 3000
cgtccctgcc gccgcccccg ccttcccccc caaaaagatt cttcttcatg cagaacatac 3060
ggcagtcaac aaagggagac ctgggtccag gtgtccaagt cacttatttc gagtaaatta 3120
gcaatgaaag aatgccatgg aatccctgcc caaatacctc tgcttatgat attgtagaat 3180
ttgatataga gttgtatccc atttaaggag taggatgtag taggaaagta ctaaaaacaa 3240
acacacaaac agaaaaccct ctttgctttg taaggtggtt cctaagataa tgtcagtgca 3300
atgctggaaa taatatttaa tatgtgaagg ttttaggctg tgttttcccc tcctgttctt 3360
tttttctgcc agccctttgt cattttgca ggtcaatgaa tcatgtagaa agagacagga 3420
gatgaaacta gaaccagtcc attttgcccc ttttttttatt ttctggtttt ggtaaaagat 3480
acaatgaggt aggaggttga gatttataaa tgaagtttaa taagtttctg tagctttgat 3540
ttttctcttt catatttgtt atcttgcata agccagaatt ggcctgtaaa atctacatat 3600
ggatattgaa gtctaaatct gttcaactag cttacactag atggagatat tttcatattc 3660
agatacactg gaatgtatga tctagccatg cgtaatatag tcaagtgttt gaaggtattt 3720
atttttaata gcgtctttag ttgtggactg gttcaagttt ttctgccaat gatttcttca 3780
aatttatcaa atatttttcc atcatgaagt aaaatgccct tgcagtcacc cttcctgaag 3840
tttgaacgac tctgctgttt taaacagttt aagcaaatgg tatatcatct tccgtttact 3900
atgtagctta actgcaggct tacgcttttg agtcagcggc caactttatt gccaccttca 3960
aaagtttatt ataatgttgt aaatttttac ttctcaaggt tagcatactt aggagttgct 4020
tcacaattag gattcaggaa agaaagaact tcagtaggaa ctgattggaa tttaatgatg 4080
cagcattcaa tgggtactaa tttcaaagaa tgatattaca gcagacacac agcagttatc 4140
ttgatttttct aggaataatt gtatgaagaa tatggctgac aacacggcct tactgccact 4200
cagcggaggc tggactaatg aacaccctac ccttctttcc tttcctctca catttcatga 4260
gcgttttgta ggtaacgaga aaattgactt gcatttgcat tacaaggagg agaaactggc 4320
aaaggggatg atggtggaag ttttgttctg tctaatgaag tgaaaaatga aaatgctaga 4380
gttttgtgca acataatagt agcagtaaaa accaagtgaa aagtctttcc aaaactgtgt 4440
taagagggca tctgctggga aacgatttga ggagaaggta ctaaattgct tggtattttc 4500
cgtaggaacc ccagagcgaa atacagtttg caaaagatgt ccagatgggt tcttctcaaa 4560
tgagacgtca tctaaagcac cctgtagaaa acacacaaat tgcagtgtct ttggtctcct 4620

gctaactcag aaaggaaatg caacacacga caacatatgt tccggaaaca gtgaatcaac 4680
tcaaaaatgt ggaataggta attacattcc aaaatacgtc tttgtacgat tttgtagtat 4740
catctctctc tctgagttga acacaaggcc tccagccaca ttcttggtca aacttacatt 4800
ttcccttttct tgaatcttaa ccagctaagg ctactctcga tgcattactg ctaaagctac 4860
cactcagaat ctctcaaaaa ctcatcttct cacagataac acctcaaagc ttgattttct 4920
ctcctttcac actgaaatca aatcttgccc ataggcaaag ggcagtgtca agtttgccac 4980
tgagatgaaa ttaggagagt ccaaactgta gaattcacgt tgtgtgttat tactttcacg 5040
aatgtctgta ttattaacta aagtatatat tggcaactaa gaagcaaagt gatataaaca 5100
tgatgacaaa ttaggccagg catggtggct tactcctata atcccaacat tttgggggc 5160
caaggtaggc agatcacttg aggtcaggat ttcaagacca gcctgaccaa catggtgaaa 5220
ccttgtctct actaaaaata caaaaattag ctgggcatgg tagcaggcac ttctagtacc 5280
agctactcag ggctgaggca ggagaatcgc ttgaacccag gagatggagg ttgcagtgag 5340
ctgagattgt accactgcac tccagtctgg gcaacagagc aagatttcat cacacacaca 5400
cacacacaca cacacacaca cattagaaat gtgtacttgg ctttgttacc tatggtatta 5460
gtgcatctat tgcatggaac ttccaagcta ctctggttgt gttaagctct tcattgggta 5520
caggtcacta gtattaagtt caggttattc ggatgcattc cacggtagtg atgacaattc 5580
atcaggctag tgtgtgtgtt caccttgtca ctcccaccac tagactaatc tcagaccttc 5640
actcaaagac acattacact aaagatgatt tgctttttg tgtttaatca agcaatggta 5700
taaaccagct tgactctccc caaacagttt ttcgtactac aaagaagttt atgaagcaga 5760
gaaatgtgaa ttgatatata tatgagattc taacccagtt ccagcattgt ttcattgtgt 5820
aattgaaatc atagacaagc cattttagcc tttgctttct tatctaaaaa aaaaaaaaaa 5880
aaaatgaagg aaggggtatt aaaaggagtg atcaaatttt aacattctct ttaattaatt 5940
cattttaat tttacttttt ttcatttatt gtgcacttac tatgtggtac tgtgctatag 6000
aggctttaac atttataaaa acactgtgaa agttgcttca gatgaatata ggtagtagaa 6060
cggcagaact agtattcaaa gccaggtctg atgaatccaa aaacaaacac ccattactcc 6120
cattttctgg gacatactta ctctacccag atgctctggg ctttgtaatg cctatgtaaa 6180
taacatagtt ttatgtttgg ttatttttcct atgtaatgtc tacttatata tctgtatcta 6240
tctcttgctt tgtttccaaa ggtaaactat gtgtctaaat gtgggcaaaa aataacacac 6300
tattccaaat tactgttcaa attcctttaa gtcagtgata attatttgtt ttgacattaa 6360
tcatgaagtt ccctgtgggt actaggtaaa cctttaatag aatgttaatg tttgtattca 6420
ttataagaat ttttggctgt tacttattta caacaatatt tcactctaat tagacattta 6480
ctaaactttc tcttgaaaac aatgcccaaa aaagaacatt agaagacacg taagctcagt 6540
tggtctctgc cactaagacc agccaacaga agcttgattt tattcaaact ttgcatttta 6600
gcatatttta tcttggaaaa ttcaattgtg ttggtttttt gtttttgttt gtattgaata 6660
gactctcaga aatccaattg ttgagtaaat cttctgggtt ttctaacctt tctttagatg 6720
ttaccctgtg tgaggaggca ttcttcaggt ttgctgttcc tacaaagttt acgcctaact 6780
ggcttagtgt cttggtagac aatttgcctg gcaccaaagt aaacgcagag agtgtagaga 6840
ggataaaacg gcaacacagc tcacaagaac agactttcca gctgctgaag ttatggaaac 6900
atcaaaacaa agaccaagat atagtcaaga agatcatcca aggtatgata atctaaaata 6960
aaaagatcaa tcagaaatca aagacaccta tttatcataa accaggaaca agactgcatg 7020
tatgtttagt tgtgtggatc ttgtttccct gttggaatca ttgttggact gaaaaagttt 7080
ccacctgata atgtagatgt gattccacaa acagttatac aaggttttgt tctcacccct 7140
gctccccagt ttccttgtaa agtatgttga acactctaag agaagagaaa tgcatttgaa 7200
ggcagggctg tatctcaggg agtcgcttcc agatccctta acgcttctgt aagcagcccc 7260
tctagaccac caaggagaag ctctataacc actttgtatc ttacattgca cctctaccaa 7320
gaagctctgt tgtatttact tggtaattct ctccaggtag gcttttcgta gcttacaaat 7380
atgttcttat taatcctcat gatatggcct gcattaaaat tattttaatg gcatatgtta 7440
tgagaattaa tgagataaaa tctgaaaagt gtttgagcct cttgtaggaa aaagctagtt 7500

acagcaaaat gttctcacat cttataagtt tatataaaga ttctccttta gaaatggtgt 7560
gagagagaaa cagagagaga tagggagaga agtgtgaaag aatctgaaga aaaggagttt 7620
catccagtgt ggactgtaag ctttacgaca catgatggaa agagttctga cttcagtaag 7680
cattgggagg acatgctaga agaaaaagga agaagagttt ccataatgca gacagggtca 7740
gtgagaaatt cattcaggtc ctcaccagta gttaaatgac tgtatagtct tgcactaccc 7800
taaaaaactt caagtatctg aaaccggggc aacagatttt aggagaccaa cgtctttgag 7860
agctgattgc ttttgcttat gcaaagagta aacttttatg ttttgagcaa accaaaagta 7920
ttctttgaac gtataattag ccctgaagcc gaaagaaaag agaaaatcag agaccgttag 7980
aattggaagc aaccaaattc cctattttat aaatgaggac attttaaccc agaaagatga 8040
accgatttgg cttagggctc acagatacta agtgactcat gtcattaata gaaatgttag 8100
ttcctccctc ttaggtttgt accctagctt attactgaaa tattctctag gctgtgtgtc 8160
tcctttagtt cctcgacctc atgtctttga gttttcagat atcctcctca tggaggtagt 8220
cctctggtgc tatgtgtatt ctttaaaggc tagttacggc aattaactta tcaactagcg 8280
cctactaatg aaactttgta ttacaaagta gctaacttga atactttcct ttttttctga 8340
aatgttatgg tggtaatttc tcaaactttt tcttagaaaa ctgagagtga tgtgtcttat 8400
tttctactgt taattttcaa aattaggagc ttcttccaaa gttttgttgg atgccaaaaa 8460
tatatagcat attatcttat tataacaaaa aatatttatc tcagttctta gaaataaatg 8520
gtgtcactta actccctctc aaaagaaaag gttatcattg aaatataatt atgaaattct 8580
gcaagaacct tttgcctcac gcttgtttta tgatggcatt ggatgaatat aaatgatgtg 8640
aacacttatc tgggcttttg ctttatgcag atattgacct ctgtgaaaac agcgtgcagc 8700
ggcacattgg acatgctaac ctcaccttcg agcagcttcg tagcttgatg gaaagcttac 8760
cgggaaagaa agtgggagca gaagacattg aaaaaacaat aaaggcatgc aaacccagtg 8820
accagatcct gaagctgctc agtttgtggc gaataaaaaa tggcgaccaa gacaccttga 8880
agggcctaat gcacgcacta aagcactcaa agacgtacca ctttcccaaa actgtcactc 8940
agagtctaaa gaagaccatc aggttccttc acagcttcac aatgtacaaa ttgtatcaga 9000
agttatttt agaaatgata ggtaaccagg tccaatcagt aaaaataagc tgcttataac 9060
tggaaatggc cattgagctg tttcctcaca attggcgaga tcccatggat gagtaaactg 9120
tttctcaggc acttgaggct ttcagtgata tctttctcat taccagtgac taattttgcc 9180
acagggtact aaaagaaact atgatgtgga gaaaggacta acatctcctc caataaaccc 9240
caaatggtta atccaactgt cagatctgga tcgttatcta ctgactatat tttcccttat 9300
tactgcttgc agtaattcaa ctggaaatta aaaaaaaaaa actagactcc actgggcctt 9360
actaaatatg ggaatgtcta acttaaatag ctttgggatt ccagctatgc tagaggcttt 9420
tattagaaag ccatattttt ttctgtaaaa gttactaata tatctgtaac actattacag 9480
tattgctatt tatattcatt cagatataag atttggacat attatcatcc tataaagaaa 9540
cggtatgact taattttaga aagaaaatta tattctgttt attatgacaa atgaaagaga 9600
aaatatatat ttttaatgga aagtttgtag cattttcta ataggtactg ccatattttt 9660
ctgtgtggag tatttttata attttatctg tataagctgt aatatcattt tatagaaaat 9720
gcattattta gtcaattgtt taatgttgga aaacatatga aatataaatt atctgaatat 9780
tagatgctct gagaaattga atgtacctta tttaaaagat tttatggttt tataactata 9840
taaatgacat tattaaagtt ttcaaatta 9869

<210> 3
<211> 1206
<212> DNA
<213> Homo sapiens

<400> 3

atgaacaagt tgctgtgctg cgcgctcgtg tttctggaca tctccattaa gtggaccacc 60
caggaaacgt ttcctccaaa gtaccttcat tatgacgaag aaacctctca tcagctgttg 120
tgtgacaaat gtcctcctgg tacctaccta aaacaacact gtacagcaaa gtggaagacc 180
gtgtgcgccc cttgccctga ccactactac acagacagct ggcacaccag tgacgagtgt 240
ctatactgca gccccgtgtg caaggagctg cagtacgtca agcaggagtg caatcgcacc 300
cacaaccgcg tgtgcgaatg caaggaaggg cgctaccttg agatagagtt ctgcttgaaa 360
cataggagct gcccctcctgg atttggagtg gtgcaagctg gaaccccaga gcgaaataca 420
gtttgcaaaa gatgtccaga tgggttcttc tcaaatgaga cgtcatctaa agcaccctgt 480
agaaaacaca caaattgcag tgtctttggt ctcctgctaa ctcagaaagg aaatgcaaca 540
cacgacaaca tatgttccgg aaacagtgaa tcaactcaaa aatgtggaat agatgttacc 600
ctgtgtgagg aggcattctt caggtttgct gttcctacaa agtttacgcc taactggctt 660
agtgtcttgg tagacaattt gcctggcacc aaagtaaacg cagagagtgt agagaggata 720
aaacggcaac acagctcaca agaacagact ttccagctgc tgaagttatg gaaacatcaa 780
aacaaagacc aagatatagt caagaagatc atccaagata ttgacctctg tgaaaacagc 840
gtgcagcggc acattggaca tgctaacctc accttcgagc agcttcgtag cttgatggaa 900
agcttaccgg gaaagaaagt gggagcagaa gacattgaaa aaacaataaa ggcatgcaaa 960
cccagtgacc agatcctgaa gctgctcagt ttgtggcgaa taaaaaatgg cgaccaagac 1020
accttgaagg gcctaatgca cgcactaaag cactcaaaga cgtaccactt tcccaaaact 1080
gtcactcaga gtctaaagaa gaccatcagg ttccttcaca gcttcacaat gtacaaattg 1140
tatcagaagt tatttttaga aatgataggt aaccaggtcc aatcagtaaa aataagctgc 1200
ttataa 1206

<210> 4
<211> 401
<212> PRT
<213> Homo sapiens

<400> 4
Met Asn Lys Leu Leu Cys Cys Ala Leu Val Phe Leu Asp Ile Ser Ile
1 5 10 15

Lys Trp Thr Thr Gln Glu Thr Phe Pro Pro Lys Tyr Leu His Tyr Asp
20 25 30

Glu Glu Thr Ser His Gln Leu Leu Cys Asp Lys Cys Pro Pro Gly Thr
35 40 45

Tyr Leu Lys Gln His Cys Thr Ala Lys Trp Lys Thr Val Cys Ala Pro
50 55 60

Cys Pro Asp His Tyr Tyr Thr Asp Ser Trp His Thr Ser Asp Glu Cys
65 70 75 80

Leu Tyr Cys Ser Pro Val Cys Lys Glu Leu Gln Tyr Val Lys Gln Glu
85 90 95

Cys Asn Arg Thr His Asn Arg Val Cys Glu Cys Lys Glu Gly Arg Tyr

```
            100                 105                 110

Leu Glu Ile Glu Phe Cys Leu Lys His Arg Ser Cys Pro Pro Gly Phe
        115                 120                 125

Gly Val Val Gln Ala Gly Thr Pro Glu Arg Asn Thr Val Cys Lys Arg
    130                 135                 140

Cys Pro Asp Gly Phe Phe Ser Asn Glu Thr Ser Ser Lys Ala Pro Cys
145                 150                 155                 160

Arg Lys His Thr Asn Cys Ser Val Phe Gly Leu Leu Leu Thr Gln Lys
                165                 170                 175

Gly Asn Ala Thr His Asp Asn Ile Cys Ser Gly Asn Ser Glu Ser Thr
            180                 185                 190

Gln Lys Cys Gly Ile Asp Val Thr Leu Cys Glu Glu Ala Phe Phe Arg
        195                 200                 205

Phe Ala Val Pro Thr Lys Phe Thr Pro Asn Trp Leu Ser Val Leu Val
    210                 215                 220

Asp Asn Leu Pro Gly Thr Lys Val Asn Ala Glu Ser Val Glu Arg Ile
225                 230                 235                 240

Lys Arg Gln His Ser Ser Gln Glu Gln Thr Phe Gln Leu Leu Lys Leu
                245                 250                 255

Trp Lys His Gln Asn Lys Asp Gln Asp Ile Val Lys Lys Ile Ile Gln
            260                 265                 270

Asp Ile Asp Leu Cys Glu Asn Ser Val Gln Arg His Ile Gly His Ala
        275                 280                 285

Asn Leu Thr Phe Glu Gln Leu Arg Ser Leu Met Glu Ser Leu Pro Gly
    290                 295                 300

Lys Lys Val Gly Ala Glu Asp Ile Glu Lys Thr Ile Lys Ala Cys Lys
305                 310                 315                 320

Pro Ser Asp Gln Ile Leu Lys Leu Leu Ser Leu Trp Arg Ile Lys Asn
                325                 330                 335

Gly Asp Gln Asp Thr Leu Lys Gly Leu Met His Ala Leu Lys His Ser
            340                 345                 350

Lys Thr Tyr His Phe Pro Lys Thr Val Thr Gln Ser Leu Lys Lys Thr
```

355 360 365

Ile Arg Phe Leu His Ser Ph Thr Met Tyr Lys Leu Tyr Gln Lys Leu
370 375 380

Phe Leu Glu Met Ile Gly Asn Gln Val Gln Ser Val Lys Ile Ser Cys
385 390 395 400

Leu

<210> 5
<211> 15
<212> DNA
<213> Homo sapiens

<400> 5
acttgaagat gaatg 15

<210> 6
<211> 15
<212> DNA
<213> Homo sapiens

<400> 6
acttgaaaat gaatg 15

<210> 7
<211> 15
<212> DNA
<213> Homo sapiens

<400> 7
gatcttggct ggatc 15

<210> 8
<211> 15
<212> DNA
<213> Homo sapiens

<400> 8
gatcttgtct ggatc 15

<210> 9
<211> 15
<212> DNA
<213> Homo sapiens

<400> 9
ccaccgcccc acccc 15

<210> 10
<211> 15
<212> DNA
<213> Homo sapiens

<400> 10
ccaccgctcc acccc 15

<210> 11
<211> 15
<212> DNA
<213> Homo sapiens

<400> 11
tccctggggg atcct 15

<210> 12
<211> 15
<212> DNA
<213> Homo sapiens

<400> 12
tccctggtgg atcct 15

<210> 13
<211> 15
<212> DNA
<213> Homo sapiens

<400> 13
gcgttaatcc tggag 15

<210> 14
<211> 15
<212> DNA

<213> Homo sapiens

<400> 14
gcgttaaccc tggag 15

<210> 15
<211> 15
<212> DNA
<213> Homo sapiens

<400> 15
cctgggccag ccgac 15

<210> 16
<211> 15
<212> DNA
<213> Homo sapiens

<400> 16
cctgggctag ccgac 15

<210> 17
<211> 15
<212> DNA
<213> Homo sapiens

<400> 17
gggagaaggc tccac 15

<210> 18
<211> 15
<212> DNA
<213> Homo sapiens

<400> 18
gggagaaagc tccac 15

<210> 19
<211> 15
<212> DNA
<213> Homo sapiens

<400> 19

ccttttacgc tgcaa 15

<210> 20
<211> 15
<212> DNA
<213> Homo sapiens

<400> 20
cctttaagc tgcaa 15

<210> 21
<211> 15
<212> DNA
<213> Homo sapiens

<400> 21
gctggtacgt gtcaa 15

<210> 22
<211> 15
<212> DNA
<213> Homo sapiens

<400> 22
gctggtatgt gtcaa 15

<210> 23
<211> 15
<212> DNA
<213> Homo sapiens

<400> 23
aggaccattg ctcag 15

<210> 24
<211> 15
<212> DNA
<213> Homo sapiens

<400> 24
aggaccactg ctcag 15

<210> 25
<211> 15
<212> DNA
<213> Homo sapiens

<400> 25
aacataatag tagca 15

<210> 26
<211> 15
<212> DNA
<213> Homo sapiens

<400> 26
aacataacag tagca 15

<210> 27
<211> 15
<212> DNA
<213> Homo sapiens

<400> 27
tattttccgt aggaa 15

<210> 28
<211> 15
<212> DNA
<213> Homo sapiens

<400> 28
tattttctgt aggaa 15

<210> 29
<211> 15
<212> DNA
<213> Homo sapiens

<400> 29
cattttagca tattt 15

<210> 30
<211> 15
<212> DNA

<213> Homo sapiens

<400> 30
cattttaaca tattt 15

<210> 31
<211> 15
<212> DNA
<213> Homo sapiens

<400> 31
aagtaaacgc agaga 15

<210> 32
<211> 15
<212> DNA
<213> Homo sapiens

<400> 32
aagtaaatgc agaga 15

<210> 33
<211> 15
<212> DNA
<213> Homo sapiens

<400> 33
gtgtagagag gataa 15

<210> 34
<211> 15
<212> DNA
<213> Homo sapiens

<400> 34
gtgtagaaag gataa 15

<210> 35
<211> 15
<212> DNA
<213> Homo sapiens

<400> 35

agaggataaa acggc 15

<210> 36
<211> 15
<212> DNA
<213> Homo sapiens

<400> 36
agaggatgaa acggc 15

<210> 37
<211> 15
<212> DNA
<213> Homo sapiens

<400> 37
tgaagttatg gaaac 15

<210> 38
<211> 15
<212> DNA
<213> Homo sapiens

<400> 38
tgaagttgtg gaaac 15

<210> 39
<211> 15
<212> DNA
<213> Homo sapiens

<400> 39
tgaagttatg gaaac 15

<210> 40
<211> 15
<212> DNA
<213> Homo sapiens

<400> 40
tgaagttgtg gaaac 15

<210> 41
<211> 15
<212> DNA
<213> Homo sapiens

<400> 41
gtatgataat ctaaa 15

<210> 42
<211> 15
<212> DNA
<213> Homo sapiens

<400> 42
gtatgatcat ctaaa 15

<210> 43
<211> 15
<212> DNA
<213> Homo sapiens

<400> 43
tagttacggc aatta 15

<210> 44
<211> 15
<212> DNA
<213> Homo sapiens

<400> 44
tagttacagc aatta 15

<210> 45
<211> 15
<212> DNA
<213> Homo sapiens

<400> 45
agagtgatgt gtctt 15

<210> 46
<211> 15
<212> DNA

<213> Homo sapiens

<400> 46
agagtgacgt gtctt 15

<210> 47
<211> 16
<212> DNA
<213> Homo sapiens

<400> 47
ggcattggat gaatat 16

<210> 48
<211> 16
<212> DNA
<213> Homo sapiens

<400> 48
ggcattgtat gaatat 16

<210> 49
<211> 15
<212> DNA
<213> Homo sapiens

<400> 49
aaacagcgtg cagcg 15

<210> 50
<211> 15
<212> DNA
<213> Homo sapiens

<400> 50
aaacagcatg cagcg 15

<210> 51
<211> 15
<212> DNA
<213> Homo sapiens

<400> 51

aaagaagacc atcag 15

<210> 52
<211> 15
<212> DNA
<213> Homo sapiens

<400> 52
aaagaagtcc atcag 15

<210> 53
<211> 15
<212> DNA
<213> Homo sapiens

<400> 53
tcagaagtta ttttt 15

<210> 54
<211> 15
<212> DNA
<213> Homo sapiens

<400> 54
tcagaagcta ttttt 15

<210> 55
<211> 15
<212> DNA
<213> Homo sapiens

<400> 55
atgtaaactt gaaga 15

<210> 56
<211> 17
<212> DNA
<213> Homo sapiens

<400> 56
ttcgcaatca ttcatct 17

<210> 57
<211> 15
<212> DNA
<213> Homo sapiens

<400> 57
atgtaaactt gaaaa 15

<210> 58
<211> 17
<212> DNA
<213> Homo sapiens

<400> 58
ttcgcaatca ttcattt 17

<210> 59
<211> 15
<212> DNA
<213> Homo sapiens

<400> 59
cgtccggatc ttggc 15

<210> 60
<211> 15
<212> DNA
<213> Homo sapiens

<400> 60
gagtccgatc cagcc 15

<210> 61
<211> 15
<212> DNA
<213> Homo sapiens

<400> 61
cgtccggatc ttgtc 15

<210> 62
<211> 15
<212> DNA

<213> Homo sapiens

<400> 62
gagtccgatc cagac 15

<210> 63
<211> 15
<212> DNA
<213> Homo sapiens

<400> 63
cagacaccac cgccc 15

<210> 64
<211> 15
<212> DNA
<213> Homo sapiens

<400> 64
gcgtgagggg tgggg 15

<210> 65
<211> 15
<212> DNA
<213> Homo sapiens

<400> 65
cagacaccac cgctc 15

<210> 66
<211> 15
<212> DNA
<213> Homo sapiens

<400> 66
gcgtgagggg tggag 15

<210> 67
<211> 15
<212> DNA
<213> Homo sapiens

<400> 67

cccacctccc tgggg 15

<210> 68
<211> 15
<212> DNA
<213> Homo sapiens

<400> 68
gcggaaagga tcccc 15

<210> 69
<211> 15
<212> DNA
<213> Homo sapiens

<400> 69
cccacctccc tggtg 15

<210> 70
<211> 15
<212> DNA
<213> Homo sapiens

<400> 70
gcggaaagga tccac 15

<210> 71
<211> 15
<212> DNA
<213> Homo sapiens

<400> 71
ctgaaagcgt taatc 15

<210> 72
<211> 15
<212> DNA
<213> Homo sapiens

<400> 72
agaaagctcc aggat 15

<210> 73
<211> 15
<212> DNA
<213> Homo sapiens

<400> 73
ctgaaagcgt taacc 15

<210> 74
<211> 15
<212> DNA
<213> Homo sapiens

<400> 74
agaaagctcc agggt 15

<210> 75
<211> 15
<212> DNA
<213> Homo sapiens

<400> 75
taagtccctg ggcca 15

<210> 76
<211> 15
<212> DNA
<213> Homo sapiens

<400> 76
gcacccgtcg gctgg 15

<210> 77
<211> 15
<212> DNA
<213> Homo sapiens

<400> 77
taagtccctg ggcta 15

<210> 78
<211> 15
<212> DNA

<213> Homo sapiens

<400> 78
gcacccgtcg gctag 15

<210> 79
<211> 15
<212> DNA
<213> Homo sapiens

<400> 79
ccggcgggga gaagg 15

<210> 80
<211> 15
<212> DNA
<213> Homo sapiens

<400> 80
gagcgagtgg agcct 15

<210> 81
<211> 15
<212> DNA
<213> Homo sapiens

<400> 81
ccggcgggga gaaag 15

<210> 82
<211> 15
<212> DNA
<213> Homo sapiens

<400> 82
gagcgagtgg agctt 15

<210> 83
<211> 15
<212> DNA
<213> Homo sapiens

<400> 83

gggtgtcctt ttacg 15

<210> 84
<211> 15
<212> DNA
<213> Homo sapiens

<400> 84
ggaactttgc agcgt 15

<210> 85
<211> 15
<212> DNA
<213> Homo sapiens

<400> 85
gggtgtcctt ttaag 15

<210> 86
<211> 15
<212> DNA
<213> Homo sapiens

<400> 86
ggaactttgc agctt 15

<210> 87
<211> 15
<212> DNA
<213> Homo sapiens

<400> 87
gtgcaagctg gtacg 15

<210> 88
<211> 15
<212> DNA
<213> Homo sapiens

<400> 88
tgcacattga cacgt 15

<210> 89
<211> 15
<212> DNA
<213> Homo sapiens

<400> 89
gtgcaagctg gtatg 15

<210> 90
<211> 15
<212> DNA
<213> Homo sapiens

<400> 90
tgcacattga cacat 15

<210> 91
<211> 15
<212> DNA
<213> Homo sapiens

<400> 91
ttccaaagga ccatt 15

<210> 92
<211> 15
<212> DNA
<213> Homo sapiens

<400> 92
attcctctga gcaat 15

<210> 93
<211> 15
<212> DNA
<213> Homo sapiens

<400> 93
ttccaaagga ccact 15

<210> 94
<211> 15
<212> DNA

<213> Homo sapiens

<400> 94
attcctctga gcagt 15

<210> 95
<211> 15
<212> DNA
<213> Homo sapiens

<400> 95
ttgtgcaaca taata 15

<210> 96
<211> 15
<212> DNA
<213> Homo sapiens

<400> 96
ttttactgct actat 15

<210> 97
<211> 15
<212> DNA
<213> Homo sapiens

<400> 97
ttgtgcaaca taaca 15

<210> 98
<211> 15
<212> DNA
<213> Homo sapiens

<400> 98
ttttactgct actgt 15

<210> 99
<211> 15
<212> DNA
<213> Homo sapiens

<400> 99

gcttggtatt ttccg 15

<210> 100
<211> 15
<212> DNA
<213> Homo sapiens

<400> 100
ctggggttcc tacgg 15

<210> 101
<211> 15
<212> DNA
<213> Homo sapiens

<400> 101
gcttggtatt ttctg 15

<210> 102
<211> 15
<212> DNA
<213> Homo sapiens

<400> 102
ctggggttcc tacag 15

<210> 103
<211> 15
<212> DNA
<213> Homo sapiens

<400> 103
actttgcatt ttagc 15

<210> 104
<211> 15
<212> DNA
<213> Homo sapiens

<400> 104
aagataaaat atgct 15

<210> 105
<211> 15
<212> DNA
<213> Homo sapiens

<400> 105
actttgcatt ttaac 15

<210> 106
<211> 15
<212> DNA
<213> Homo sapiens

<400> 106
aagataaaat atgtt 15

<210> 107
<211> 15
<212> DNA
<213> Homo sapiens

<400> 107
gcaccaaagt aaacg 15

<210> 108
<211> 15
<212> DNA
<213> Homo sapiens

<400> 108
ctacactctc tgcgt 15

<210> 109
<211> 15
<212> DNA
<213> Homo sapiens

<400> 109
gcaccaaagt aaatg 15

<210> 110
<211> 15
<212> DNA

<213> Homo sapiens

<400> 110
ctacactctc tgcat 15

<210> 111
<211> 15
<212> DNA
<213> Homo sapiens

<400> 111
cagagagtgt agaga 15

<210> 112
<211> 13
<212> DNA
<213> Homo sapiens

<400> 112
cgcgttttat cct 13

<210> 113
<211> 15
<212> DNA
<213> Homo sapiens

<400> 113
cagagagtgt agaaa 15

<210> 114
<211> 13
<212> DNA
<213> Homo sapiens

<400> 114
cgcgttttat ctt 13

<210> 115
<211> 15
<212> DNA
<213> Homo sapiens

<400> 115

gtgtagagag gataa 15

<210> 116
<211> 15
<212> DNA
<213> Homo sapiens

<400> 116
tgtgttgccg tttta 15

<210> 117
<211> 15
<212> DNA
<213> Homo sapiens

<400> 117
gtgtagagag gatga 15

<210> 118
<211> 15
<212> DNA
<213> Homo sapiens

<400> 118
tgtgttgccg tttca 15

<210> 119
<211> 15
<212> DNA
<213> Homo sapiens

<400> 119
agctgctgaa gttat 15

<210> 120
<211> 15
<212> DNA
<213> Homo sapiens

<400> 120
tttgatgttt ccata 15

<210> 121
<211> 15
<212> DNA
<213> Homo sapiens

<400> 121
agctgctgaa gttgt 15

<210> 122
<211> 15
<212> DNA
<213> Homo sapiens

<400> 122
tttgatgttt ccaca 15

<210> 123
<211> 15
<212> DNA
<213> Homo sapiens

<400> 123
tccaaggtat gataa 15

<210> 124
<211> 15
<212> DNA
<213> Homo sapiens

<400> 124
ttttatttta gatta 15

<210> 125
<211> 15
<212> DNA
<213> Homo sapiens

<400> 125
tccaaggtat gatca 15

<210> 126
<211> 15
<212> DNA

<213> Homo sapiens

<400> 126
ttttatttta gatga 15

<210> 127
<211> 15
<212> DNA
<213> Homo sapiens

<400> 127
aaaggctagt tacgg 15

<210> 128
<211> 15
<212> DNA
<213> Homo sapiens

<400> 128
tgataagtta attgc 15

<210> 129
<211> 15
<212> DNA
<213> Homo sapiens

<400> 129
aaaggctagt tacag 15

<210> 130
<211> 15
<212> DNA
<213> Homo sapiens

<400> 130
tgataagtta atttc 15

<210> 131
<211> 15
<212> DNA
<213> Homo sapiens

<400> 131

aaactgagag tgatg 15

<210> 132
<211> 15
<212> DNA
<213> Homo sapiens

<400> 132
gaaaataaga cacat 15

<210> 133
<211> 15
<212> DNA
<213> Homo sapiens

<400> 133
aaactgagag tgacg 15

<210> 134
<211> 15
<212> DNA
<213> Homo sapiens

<400> 134
gaaaataaga cacgt 15

<210> 135
<211> 15
<212> DNA
<213> Homo sapiens

<400> 135
tatgatggca ttgga 15

<210> 136
<211> 15
<212> DNA
<213> Homo sapiens

<400> 136
catttatatt catcc 15

<210> 137
<211> 15
<212> DNA
<213> Homo sapiens

<400> 137
tatgatggca ttgta 15

<210> 138
<211> 15
<212> DNA
<213> Homo sapiens

<400> 138
catttatatt catac 15

<210> 139
<211> 15
<212> DNA
<213> Homo sapiens

<400> 139
ctgtgaaaac agcgt 15

<210> 140
<211> 15
<212> DNA
<213> Homo sapiens

<400> 140
atgtgccgct gcacg 15

<210> 141
<211> 15
<212> DNA
<213> Homo sapiens

<400> 141
ctgtgaaaac agcat 15

<210> 142
<211> 15
<212> DNA

<213> Homo sapiens

<400> 142
atgtgccgct gcatg 15

<210> 143
<211> 15
<212> DNA
<213> Homo sapiens

<400> 143
gagtctaaag aagac 15

<210> 144
<211> 15
<212> DNA
<213> Homo sapiens

<400> 144
aggaacctga tggtc 15

<210> 145
<211> 15
<212> DNA
<213> Homo sapiens

<400> 145
gagtctaaag aagtc 15

<210> 146
<211> 15
<212> DNA
<213> Homo sapiens

<400> 146
aggaacctga tggac 15

<210> 147
<211> 15
<212> DNA
<213> Homo sapiens

<400> 147

attgtatcag aagtt 15

<210> 148
<211> 15
<212> DNA
<213> Homo sapiens

<400> 148
atttctaaaa ataac 15

<210> 149
<211> 15
<212> DNA
<213> Homo sapiens

<400> 149
attgtatcag aagct 15

<210> 150
<211> 15
<212> DNA
<213> Homo sapiens

<400> 150
atttctaaaa atagc 15

<210> 151
<211> 10
<212> DNA
<213> Homo sapiens

<400> 151
taaacttgaa 10

<210> 152
<211> 10
<212> DNA
<213> Homo sapiens

<400> 152
aatcattcat 10

<210> 153
<211> 10
<212> DNA
<213> Homo sapiens

<400> 153
ccggatcttg 10

<210> 154
<211> 10
<212> DNA
<213> Homo sapiens

<400> 154
tccgatccag 10

<210> 155
<211> 10
<212> DNA
<213> Homo sapiens

<400> 155
acaccaccgc 10

<210> 156
<211> 10
<212> DNA
<213> Homo sapiens

<400> 156
tgaggggtgg 10

<210> 157
<211> 10
<212> DNA
<213> Homo sapiens

<400> 157
acctccctgg 10

<210> 158
<211> 10
<212> DNA

<213> Homo sapiens

<400> 158
gaaaggatcc 10

<210> 159
<211> 10
<212> DNA
<213> Homo sapiens

<400> 159
aaagcgttaa 10

<210> 160
<211> 10
<212> DNA
<213> Homo sapiens

<400> 160
aagctccagg 10

<210> 161
<211> 10
<212> DNA
<213> Homo sapiens

<400> 161
gtccctgggc 10

<210> 162
<211> 10
<212> DNA
<213> Homo sapiens

<400> 162
cccgtcggct 10

<210> 163
<211> 10
<212> DNA
<213> Homo sapiens

<400> 163

gcggggagaa 10

<210> 164
<211> 10
<212> DNA
<213> Homo sapiens

<400> 164
cgagtggagc 10

<210> 165
<211> 10
<212> DNA
<213> Homo sapiens

<400> 165
tgtcctttta 10

<210> 166
<211> 10
<212> DNA
<213> Homo sapiens

<400> 166
actttgcagc 10

<210> 167
<211> 10
<212> DNA
<213> Homo sapiens

<400> 167
caagctggta 10

<210> 168
<211> 10
<212> DNA
<213> Homo sapiens

<400> 168
acattgacac 10

<210> 169
<211> 10
<212> DNA
<213> Homo sapiens

<400> 169
caaaggacca 10

<210> 170
<211> 10
<212> DNA
<213> Homo sapiens

<400> 170
cctctgagca 10

<210> 171
<211> 10
<212> DNA
<213> Homo sapiens

<400> 171
tgcaacataa 10

<210> 172
<211> 10
<212> DNA
<213> Homo sapiens

<400> 172
tactgctact 10

<210> 173
<211> 10
<212> DNA
<213> Homo sapiens

<400> 173
tggtattttc 10

<210> 174
<211> 10
<212> DNA

<213> Homo sapiens

<400> 174
gggttcctac 10

<210> 175
<211> 10
<212> DNA
<213> Homo sapiens

<400> 175
ttgcatttta 10

<210> 176
<211> 10
<212> DNA
<213> Homo sapiens

<400> 176
ataaaatatg 10

<210> 177
<211> 10
<212> DNA
<213> Homo sapiens

<400> 177
ccaaagtaaa 10

<210> 178
<211> 10
<212> DNA
<213> Homo sapiens

<400> 178
cactctctgc 10

<210> 179
<211> 10
<212> DNA
<213> Homo sapiens

<400> 179

agagtgtaga 10

<210> 180
<211> 10
<212> DNA
<213> Homo sapiens

<400> 180
gttttatcct 10

<210> 181
<211> 10
<212> DNA
<213> Homo sapiens

<400> 181
tagagaggat 10

<210> 182
<211> 10
<212> DNA
<213> Homo sapiens

<400> 182
gttgccgttt 10

<210> 183
<211> 10
<212> DNA
<213> Homo sapiens

<400> 183
tgctgaagtt 10

<210> 184
<211> 10
<212> DNA
<213> Homo sapiens

<400> 184
gatgtttcca 10

<210> 185
<211> 10
<212> DNA
<213> Homo sapiens

<400> 185
aaggtatgat 10

<210> 186
<211> 10
<212> DNA
<213> Homo sapiens

<400> 186
tattttagat 10

<210> 187
<211> 10
<212> DNA
<213> Homo sapiens

<400> 187
ggctagttac 10

<210> 188
<211> 10
<212> DNA
<213> Homo sapiens

<400> 188
agttaattgc 10

<210> 189
<211> 10
<212> DNA
<213> Homo sapiens

<400> 189
ctgagagtga 10

<210> 190
<211> 10
<212> DNA

<213> Homo sapiens

<400> 190
aataagacac 10

<210> 191
<211> 9
<212> DNA
<213> Homo sapiens

<400> 191
atggcattg 9

<210> 192
<211> 10
<212> DNA
<213> Homo sapiens

<400> 192
ttatattcat 10

<210> 193
<211> 10
<212> DNA
<213> Homo sapiens

<400> 193
tgaaaacagc 10

<210> 194
<211> 10
<212> DNA
<213> Homo sapiens

<400> 194
tgccgctgca 10

<210> 195
<211> 10
<212> DNA
<213> Homo sapiens

<400> 195

tctaaagaag 10

<210> 196
<211> 10
<212> DNA
<213> Homo sapiens

<400> 196
aacctgatgg 10

<210> 197
<211> 10
<212> DNA
<213> Homo sapiens

<400> 197
gtatcagaag 10

<210> 198
<211> 10
<212> DNA
<213> Homo sapiens

<400> 198
tctaaaaata 10

<210> 199
<211> 20
<212> DNA
<213> Homo sapiens

<400> 199
gtacggcgga aactcacagc 20

<210> 200
<211> 23
<212> DNA
<213> Homo sapiens

<400> 200
gcacagcaac ttgttcattg tgg 23

<210> 201
<211> 20
<212> DNA
<213> Homo sapiens

<400> 201
tctcccaggg gacagacacc 20

<210> 202
<211> 23
<212> DNA
<213> Homo sapiens

<400> 202
taaaaggaca ccctagggga agc 23

<210> 203
<211> 22
<212> DNA
<213> Homo sapiens

<400> 203
atcctggagc tttctgcaca cc 22

<210> 204
<211> 26
<212> DNA
<213> Homo sapiens

<400> 204
ctgaatctaa gggaccactt ctttgc 26

<210> 205
<211> 24
<212> DNA
<213> Homo sapiens

<400> 205
gctaagatga tgccactgtg ttcc 24

<210> 206
<211> 23
<212> DNA

<213> Homo sapiens

<400> 206
gccctgtagt ggcaaagtat tcc 23

<210> 207
<211> 24
<212> DNA
<213> Homo sapiens

<400> 207
cctttcctct cacatttcat gagc 24

<210> 208
<211> 22
<212> DNA
<213> Homo sapiens

<400> 208
caaacttgac actgcccttt gc 22

<210> 209
<211> 25
<212> DNA
<213> Homo sapiens

<400> 209
acacgtaagc tcagttggtc tctgc 25

<210> 210
<211> 22
<212> DNA
<213> Homo sapiens

<400> 210
gagcaggggt gagaacaaaa cc 22

<210> 211
<211> 23
<212> DNA
<213> Homo sapiens

<400> 211

cttccaaagt tttgttggat gcc 23

<210> 212
<211> 24
<212> DNA
<213> Homo sapiens

<400> 212
agtttactca tccatgggat ctcg 24

<210> 213
<211> 20
<212> DNA
<213> Homo sapiens

<400> 213
cccagcgaga ggacaaaggt 20

<210> 214
<211> 19
<212> DNA
<213> Homo sapiens

<400> 214
ggaaacctca ggggcttgg 19

<210> 215
<211> 21
<212> DNA
<213> Homo sapiens

<400> 215
cccagccctg aaagcgttaa t 21

<210> 216
<211> 20
<212> DNA
<213> Homo sapiens

<400> 216
aaagcggttt cctgctccag 20

<210> 217
<211> 20
<212> DNA
<213> Homo sapiens

<400> 217
ccgcccaagc ttcctaaaaa 20

<210> 218
<211> 22
<212> DNA
<213> Homo sapiens

<400> 218
tctccctctc tctcgctgtc tg 22

<210> 219
<211> 20
<212> DNA
<213> Homo sapiens

<400> 219
agtggaccac ccaggaaacg 20

<210> 220
<211> 20
<212> DNA
<213> Homo sapiens

<400> 220
gcaatggtcc tttggaagca 20

<210> 221
<211> 20
<212> DNA
<213> Homo sapiens

<400> 221
aactggcaaa ggggatgatg 20

<210> 222
<211> 22
<212> DNA

<213> Homo sapiens

<400> 222
gcatcgagag tagccttagc tg 22

<210> 223
<211> 20
<212> DNA
<213> Homo sapiens

<400> 223
accagccaac agaagcttga 20

<210> 224
<211> 22
<212> DNA
<213> Homo sapiens

<400> 224
gtccaacaat gattccaaca gg 22

<210> 225
<211> 22
<212> DNA
<213> Homo sapiens

<400> 225
ggtgtcactt aactccctct ca 22

<210> 226
<211> 20
<212> DNA
<213> Homo sapiens

<400> 226
ctgattggac ctggttacct 20

<210> 227
<211> 26
<212> DNA
<213> Homo sapiens

<400> 227

ctgtaaacaa tttcagtggc aacccg 26

<210> 228
<211> 26
<212> DNA
<213> Homo sapiens

<400> 228
ccgtgctatt ctgcattcac tccttg 26

<210> 229
<211> 26
<212> DNA
<213> Homo sapiens

<400> 229
cgtaggaagc tccgatacca atagcc 26

<210> 230
<211> 23
<212> DNA
<213> Homo sapiens

<400> 230
ggagagcagg ggaaaaaaaa gcc 23

<210> 231
<211> 26
<212> DNA
<213> Homo sapiens

<400> 231
ttgaggtttc agaacccgaa gtgaag 26

<210> 232
<211> 25
<212> DNA
<213> Homo sapiens

<400> 232
caacaggaag tatcgcctgc ctttg 25

<210> 233
<211> 26
<212> DNA
<213> Homo sapiens

<400> 233
gttaatcctg gagctttctg cacacc 26

<210> 234
<211> 26
<212> DNA
<213> Homo sapiens

<400> 234
tcgctgtctg tctctctctt gctgtc 26

<210> 235
<211> 24
<212> DNA
<213> Homo sapiens

<400> 235
tcatgctaag atgatgccac tgtg 24

<210> 236
<211> 24
<212> DNA
<213> Homo sapiens

<400> 236
gcaaagtatt cctctgagca atgg 24

<210> 237
<211> 23
<212> DNA
<213> Homo sapiens

<400> 237
ctgtgttaag agggcatctg ctg 23

<210> 238
<211> 22
<212> DNA

<213> Homo sapiens

<400> 238
ttgaccaaga atgtggctgg ag 22

<210> 239
<211> 25
<212> DNA
<213> Homo sapiens

<400> 239
gccactaaga ccagccaaca gaagc 25

<210> 240
<211> 25
<212> DNA
<213> Homo sapiens

<400> 240
gagcaggggt gagaacaaaa ccttg 25

<210> 241
<211> 26
<212> DNA
<213> Homo sapiens

<400> 241
ggctgtgtgt ctcctttagt tcctcg 26

<210> 242
<211> 26
<212> DNA
<213> Homo sapiens

<400> 242
ctgacagttg gattaaccat ttgggg 26

<210> 243
<211> 26
<212> DNA
<213> Homo sapiens

<400> 243

tgggaaagtg gtacgtcttt gagtgc 26

<210> 244
<211> 24
<212> DNA
<213> Homo sapiens

<400> 244
ttgacctctg tgaaaacagc gtgc 24

<210> 245
<211> 26
<212> DNA
<213> Homo sapiens

<400> 245
ctgaaagcct caagtgcctg agaaac 26

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/18803

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) : C07H 21/02, 21/04; A01K 67/00
US CL : 536/23.1; 800/8
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1; 800/8

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

STN, EAST, Sequence search of commercial databases

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CHAMBERS, T.J. Regulation of the differentiation and function of osteoclasts. Journal of Pathology September 2000. Vol. 192. pages 4-13.	1-26
Y	YASUDA, H. et al. Identity of osteoclastogenesis inhibitory factor (OCIF) and osteoprotegerin (OPG): a mechanism by which OPG/OCIF inhibits osteoclastogenesis in vitro. Endcrinology, March 1998. Vol. 139., No. 3. pages 1329-37.	1-26
Y	SIMONET, W.S. et al. Osteoprotgerin: a novel secreted protein involved in the regulation of bone density. 18 April 1997. Vol. 89., No. 2. pages 309-19.	1-26

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"B" earlier document published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 NOVEMBER 2000	21 DEC 2000

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	JOSEPH T. WOITACH
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196

DELLA MAE COLLINS
PARALEGAL SPECIALIST
TECHNOLOGY CENTER 1800

Form PCT/ISA/210 (second sheet) (July 1998)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/18803

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	LACEY, D.L. et al. Osteoprotegerin ligand is a cytokine that regulates osteoclast differentiation and activation. 17 April 1998. Vol 93., No. 2. pages 165-76.	1-26
Y	MORINAGA, T. et al. Cloning and characterization of the gene encoding human osteoprotegerin/osteoclastogenesis-inhibitory factor.FEBS. June 1998. Vol 254., No. 3. pages 685-91.	1-26
Y	TOMOYASU, A. et al. Characterization of monomeric and homdimeric forms of osteoclastogenesis inhibitory factor. Biochem. Biophys. Res. Commun. April 1998. Vol 245. No. 2. abstract.	1-26
Y	YAMAGUCHI, K. et al. Characterization of structural domains of human osteoclastogenesis inhibitory factor. Journal of Biological Chemistry. February 1998. Vol 273. No. 9. pages 5117-23.	1-26
Y	Database Genbank, National Center for Biotechnology Information (NCBI) on as a division of the National Library of Medicine (NLM) at the National Institutes of Health (NIH), No. U94332, May 1997.	1-26
Y	Database Genbank, National Center for Biotechnology Information (NCBI) on as a division of the National Library of Medicine (NLM) at the National Institutes of Health (NIH), No. U94331, May 1997.	1-26

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/18803

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☒ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.